Spectrophysics

Spectrophysics

ANNE P. THORNE
Senior Lecturer in Physics,
Imperial College,
University of London

SECOND EDITION

LONDON NEW YORK
CHAPMAN AND HALL

Published by
Chapman and Hall Ltd
11 New Fetter Lane, London EC4P 4EE
First published in 1974
Second edition 1988
Published in the USA by
Chapman and Hall
29 West 35th Street, New York NY 10001

© 1974, 1988 Anne P. Thorne
Reprint of the original edition 1974

J. W. Arrowsmith Ltd, Bristol

ISBN-13:978-0-412-27470-1 e-ISBN-13:978-94-009-1193-2
DOI: 10.1007/978-94-009-1193-2

This paperback edition is sold subject to the condition that it shall not, by way of trade or otherwise, be lent, resold, hired out, or otherwise circulated without the publisher's prior consent in any form of binding or cover other than that in which it is published and without a similar condition including this condition being imposed on the subsequent purchaser.

All rights reserved. No part of this book may be reprinted or reproduced, or utilized in any form or by any electronic, mechanical or other means, now known or hereafter invented, including photocopying and recording, or in any information storage and retrieval system, without permission in writing from the publisher.

British Library Cataloguing in Publication Data

Thorne, Anne P.
Spectrophysics.—2nd edition
1. Spectrum analysis
I. Title
535.8'4 QC451
ISBN-13:978-0-412-27470-1

Library of Congress Cataloging in Publication Data

Thorne, Anne P.
Spectrophysics.

Bibliography: p.
Includes index.
1. Spectrum analysis. I. Title.
QC451.T47 1988 535'.8'4 87-30004
ISBN-13:978-0-412-27470-1

To my
forbearing family
and most of all
to Peter

Cover photograph shows 10 metre Z–pinch discharge, 200 000 amps, in helium. Courtesy of Mr N. Jackson and the Spectroscopy Group at the Blackett Laboratory, Imperial College, London.

Contents

Preface to the first edition		xiii
Preface to the second edition		xv
1	*Introduction*	1
1.1	The uses of spectroscopy	1
1.2	Scope of this book	3
1.3	The electromagnetic spectrum	4
1.4	Term diagrams and line series	7
1.5	Sources, detectors and dispersers	11
1.6	Survey of spectral regions	14
1.7	Wavelength standards	15
1.8	Units used in this book	17
	References	18
2	*A Summary of atomic structure*	20
2.1	Schrödinger's equation: wave functions	20
2.2	The hydrogen atom: degeneracy and spin	23
2.3	The Pauli exclusion principle and shell structure	27
2.4	The central field model	29
2.5	Interlude: perturbation theory	31
2.6	Electrostatic interactions: the quantum numbers L and S	34
2.7	Magnetic interactions: the quantum numbers j and J	38
2.8	Radiative transitions and selection rules	43
2.9	Zeeman and Stark effects	47
2.10	Configuration interaction and auto-ionization	54
2.11	Hyperfine structure and isotope shift	58
2.12	Ionic spectra	62
2.13	Inner shell and X-ray spectra	63
	References	64

3 *A summary of molecular structure* 65
3.1 Interatomic potentials 65
3.2 The Born–Oppenheimer approximation 66
3.3 Electronic energy of diatomic molecules 67
3.4 Vibrational energy of diatomic molecules 71
3.5 Rotational energy of diatomic molecules 74
3.6 Spectra of diatomic molecules 77
3.7 Transition probabilities for diatomic molecules 84
3.8 Some further points on spectra of diatomic molecules 85
3.9 Remarks on polyatomic molecules 90
References 92

4 *Light sources and detectors* 93
4.1 Definitions of light flux 93
4.2 Blackbody radiation and sources of continuum 96
4.3 Intensity standards 99
4.4 The excitation of line spectra 100
4.5 The Einstein probability coefficients 101
4.6 The different types of light source 104
4.7 Traditional sources: flames, arcs, sparks and glow discharges 105
4.8 Special sources 107
4.9 Absorption spectroscopy 109
4.10 General remarks on detectors 113
4.11 Detectors for the infra-red 115
4.12 Detectors for the visible and ultra-violet 118
4.13 Photographic emulsions 119
4.14 Array detectors 121
References 122

5 *Dispersion and resolving power: prism spectrographs* 125
5.1 General remarks 125
5.2 Dispersion and resolving power 126
5.3 Slit width 130
5.4 Throughput and illumination 132
5.5 Prism instruments 134
5.6 Deviation and dispersion 137
5.7 Resolving power 140
5.8 Prism materials 141
5.9 Image defects 142
5.10 Merits of prism spectrometers 143
References 143

6 *Diffraction gratings* 144
6.1 Plane grating: condition for maxima 144

6.2	Intensity distribution from ideal diffraction grating	146
6.3	Dispersion and resolving power	150
6.4	Concave gratings	153
6.5	Grating mountings	156
6.6	Production and characteristics of gratings	161
6.7	Gratings for special purposes	165
	References	169
7	*Interferometers*	171
7.1	Some basic concepts of interferometric spectroscopy	171
7.2	Resolution and throughput	173
7.3	Fabry–Perot interferometers: intensity distribution	175
7.4	Fabry–Perot interferometers: resolution and limitations	180
7.5	Fabry–Perot interferometers: methods of use	183
7.6	Michelson interferometers and Fourier transform spectroscopy	185
7.7	Fourier transform spectroscopy: instrument function, resolution, sampling and aliasing	190
7.8	Fourier transform spectroscopy: computing and control	196
7.9	Fourier transform versus grating spectroscopy: a comparison	198
	References	201
8	*Laser spectroscopy*	202
8.1	Introduction to lasers	202
8.2	Types of laser	205
8.3	Optical pumping and saturation	209
8.4	Spectral and temporal resolution	212
8.5	Laser absorption and excitation experiments	215
8.6	Doppler-free spectroscopy	218
8.7	High intensity effects, cooling and trapping	221
	References	224
9	*Other spectroscopic methods: microwave, radiofrequency, level-crossing and Raman*	227
9.1	Introduction	227
9.2	Microwave spectroscopy	228
9.3	Microwave spectroscopy of gases: the principal applications	230
9.4	Magnetic resonance spectroscopy: general considerations	232
9.5	Electron spin and nuclear magnetic resonance spectroscopy (ESR and NMR)	233
9.6	Atomic and molecular beam magnetic resonance	234
9.7	Transitions in and information from beam magnetic resonance	237
9.8	Double resonance and optical pumping experiments	242
9.9	Level-crossing experiments	245

9.10	Raman spectroscopy	247
9.11	Coherent anti-Stokes Raman spectroscopy (CARS)	251
	References	253
10	*Width and shape of spectral lines*	255
10.1	Line profiles	255
10.2	Natural broadening	257
10.3	Doppler broadening	262
10.4	Introduction to pressure broadening	265
10.5	Pressure broadening: types of perturbation	268
10.6	Pressure broadening: the impact approximation	272
10.7	Pressure broadening: the quasi-static approximation	277
10.8	Experimental aspects of pressure broadening	280
10.9	Convolution of line profiles: the Voigt profile	281
	References	285
11	*Emission and absorption of line radiation*	286
11.1	Introduction	286
11.2	Boltzmann distribution of population and state sum	286
11.3	Einstein coefficients and line strength	288
11.4	Absorption coefficient	293
11.5	Oscillator strength	296
11.6	The *f*-sum rule	301
11.7	Optical depth and equivalent width	302
11.8	Curve of growth	304
11.9	Emission lines; source function and radiative transfer	308
11.10	*f*-values in diatomic molecules	311
	References	316
12	*Experimental determination of transition probabilities and radiative lifetimes*	318
12.1	Introduction	318
12.2	Remarks on the calculation of transition probabilities	320
12.3	Emission measurements	321
12.4	Absorption and dispersion measurements	322
12.5	Lifetime measurements	328
12.6	Combinations of methods	335
	References	337
13	*Elementary plasma spectroscopy*	339
13.1	Introduction	339
13.2	Debye radius	340
13.3	Plasma oscillations	344
13.4	Energy distribution and dissociation equilibrium	347

13.5 Saha's equation for ionization equilibrium 349
13.6 Depression of ionization potential and Inglis–Teller limit 351
13.7 Electron collisional rates 352
13.8 Local thermodynamic equilibrium 355
13.9 Coronal equilibrium and ionization equilibrium 358
13.10 Continuous emission and absorption 360
13.11 The continuous absorption cross-section 363
13.12 The continuous emission 366
13.13 Other continua 368
13.14 Applications of plasma spectroscopy 369
13.15 Spectroscopic measurement of temperature 370
13.16 Spectroscopic determination of electron density 376
References 377

Appendix 379
Index 381

Preface to the first edition

This book describes the methods of experimental spectroscopy and their use in the study of physical phenomena. The applications of optical spectroscopy may be grouped under three broad headings: chemical analysis, elucidation of atomic and molecular structure, and investigations of the interactions of radiating atoms and molecules with their environment. I have used the word 'Spectrophysics' for the third of these by analogy with spectrochemistry for the first and in preference to 'quantitative spectroscopy'.

A number of textbooks treat atomic and molecular structure at varying levels of profundity, but elementary spectrophysics is not, so far as I am aware, covered in any one existing book. There is moreover a lack of up-to-date books on experimental techniques that treat in a fairly elementary fashion interferometric, Fourier transform and radiofrequency methods as well as prism and grating spectroscopy. In view of the importance of spectrophysics in astrophysics and plasma physics as well as in atomic and molecular spectroscopy there seemed a place for a book describing both the experimental methods and their spectrophysical applications.

This book is directed at students on a first degree course in physics and at postgraduates beginning research in the fields of experimental atomic or molecular spectroscopy, plasma physics or astrophysics. It is hoped this book will enable such students to set their own particular research problems in the context of similar problems and to compare the various methods of tackling them. The book is not, however, intended as a research manual, for it does not contain sufficient experimental detail for this purpose. In particular, Chapter 7 on microwave and radiofrequency spectroscopy describes only the principles of the various methods and the types of problem to which they can be applied.

The rather long chapter on atomic and molecular structure, Chapter 2, has been included to explain the concepts and terminology used in the later chapters for the benefit of any student who is not familiar with the subject. It is in no sense intended as a potted course on atomic structure. Some acquaintance with elementary wave mechanics is assumed in both this and later chapters.

A few remarks on the contents of Chapters 8 to 11 should help to indicate the scope of this book. Pressure broadening of spectral lines is discussed more fully

than is usual in elementary textbooks, but from a physical rather than a mathematical point of view. Transition probabilities, line and oscillator strengths, absorption coefficients, and the relations between these quantities are treated at some length because they seem to engender considerable confusion among students. Moreover, topics such as optical depth, equivalent width, curves of growth and radiative transfer are often totally unfamiliar to students because they are usually to be found only in textbooks on astrophysics. The last chapter introduces some basic ideas of plasma physics, discusses ionization, continuous radiation and thermodynamic equilibrium, and shows how spectroscopy may be applied to problems in both laboratory and astrophysical plasmas.

This book has been completed at a time when the rapid development of tunable dye lasers is opening up new possibilities in all branches of spectroscopy. I have remarked on some of the new techniques at appropriate places in the text, but undoubtedly many more will pass into common usage in the next few years.

I owe my sincere thanks to a number of friends and colleagues who have helped me in writing this book. In particular, Dr H. G. Kuhn, FRS, of Oxford University most kindly read the whole manuscript and made numerous suggestions for its improvement. Dr R. C. M. Learner of Imperial College has provided a wealth of constructive criticism of most of the book, and Dr D. D. Burgess, also of this college, has done likewise for the remainder. Finally, I must thank my husband and children for putting up, on the whole patiently, with a constantly pre-occupied wife and mother.

Preface to the second edition

Spectrophysics has been out of print for some years, and a number of colleagues who have found it useful in their teaching and as a quick reference for some of their research have persuaded me to resurrect it. The minimal changes and additions I at first envisaged have actually turned into a substantial rewriting of the book, although it still covers the same field at the same general level.

As to the changes, one's colleagues are much better at making suggestions for what should be added than for what should be removed. The principal addition is a chapter on laser spectroscopy. I remarked in the Preface to the first edition on the possibilities then opening up with the development of tunable dye lasers, and it is unthinkable now not to include laser spectroscopy with the other experimental techniques. The difficulties have been to avoid turning the chapter into a book, and to restrict it to techniques and applications that will not be out of date almost before the book is published.

The other main structural change is the splitting of the summary of atomic and molecular structure into two chapters. The molecular chapter now contains a little more discussion of interatomic potentials as background for the subsequent treatment of pressure broadening.

Much more spectroscopy is still done by 'conventional' methods than with lasers. I have updated the material on sources and detectors, particularly the latter, and paid more attention to signal-to-noise limitations. In the three chapters on prisms, gratings and interferometers I have improved (I hope) the presentation of resolving power and throughput and have rewritten and added somewhat to the section on Fourier transform spectroscopy (largely as a result of personal experience of this technique in the last few years). Raman spectroscopy and coherent anti-Stokes Raman spectroscopy (CARS) now put in a brief appearance at the end of the chapter on the 'other techniques'; they are not entirely logical companions to radiofrequency and magnetic resonance spectroscopy, but because of their increased importance it seemed right to include some mention of them. The remaining chapters have been updated and partly rewritten, but the content has not been greatly changed.

It proved impossible, without cutting out a whole chapter, to omit as much material as I have put in, but I have eliminated parts that seemed redundant or

out of date, and the book is not very much longer than it was. I hope it will continue to serve as a readable undergraduate text and as a handy background survey in research.

I should like to thank the many colleagues who have helped me, wittingly or unwittingly, in correcting errors and bringing the book up to date – in particular, Drs K. Burnett, P. L. Knight and R. C. M. Learner of this college, Dr J. W. Brault of the National Solar Observatory, Dr P. L. Smith of Harvard College Observatory and Dr M. C. E. Huber of ETH–Zürich. Finally, thank you to my husband for putting up with it all over again.

ONE

Introduction

1.1 The uses of spectroscopy

Spectroscopy can be used to determine the identity, the structure and the environment of atoms and molecules by analysis of the radiation emitted or absorbed by them. The light from a gaseous discharge, when analysed by wavelength to form a spectrum, is found to consist of discrete lines and bands, perhaps overlying a continuum. Each line or band is characteristic of a particular atom or molecule, and once the characteristic line pattern of an atom is known its appearance in the spectrum establishes the presence of the atom in the source. This aspect of spectroscopy is known as spectrochemical analysis, and it may be made quantitative by measuring relative intensities as well as wavelengths. Secondly, one may deduce from the line or band pattern the characteristic energy levels, or stationary states, of the atom or molecule. This provides the experimental basis from which the theories of atomic and molecular structure have been developed and are still developing. Finally, the physical properties (temperature, pressure, etc.) of the gas or plasma containing the emitting or absorbing particles affect the intensity and wavelength distribution of the radiation in various ways. The study of these effects, with a view to establishing the physical parameters of the source, may be called, by analogy to spectrochemistry, spectrophysics.

For the first couple of centuries after Newton first observed a spectrum, spectroscopy was put to none of these uses. Although the wavelengths of many spectral lines were measured during the first half of the nineteenth century – including the absorption lines in the Sun's spectrum discovered by Fraunhofer and emission lines in the spectra of flames – it was not until about 1860 that Kirchhoff and Bunsen confirmed the link between particular wavelengths and particular atoms and so inaugurated spectrochemistry. In the following few years several elements were first discovered spectroscopically (rubidium, cesium, thallium and indium, for example). In addition to its laboratory applications, the new technique was immediately directed to the identification of the Fraunhofer lines in the Sun's spectrum. Helium was 'discovered' in the Sun by Lockyer in 1868, many years before it was isolated in the laboratory.

The identification of atoms and molecules in other stars, in nebulae and in interstellar space, together with estimates of their relative abundances, has continued to the present and is of the utmost importance to astrophysics and cosmology.

The spectroscopic foundations of atomic structure were laid from 1885 onwards with the attempts by Balmer, Rydberg and others to group the observed spectral lines of a given atom in some meaningful fashion. The splitting of spectral lines in a magnetic field, discovered by Zeeman in 1896, was interpreted by Lorentz in terms of an oscillating charged particle whose charge/mass ratio identified it with the electron discovered at about the same time by Thompson. However, it was not until 1913 that Bohr, using the atomic model recently proposed by Rutherford, produced the theory of the hydrogen atom that really established the link between spectra and structure. The part played by spectroscopy in atomic physics grew rapidly in scope in the 1920s and 1930s with the introduction of quantum mechanics and the discovery of electron spin and nuclear spin, and it is far from being played out now. The unravelling of the spectra of complex atoms (such as the rare earths) and of highly ionized species, the determination of transition probabilities, and the study of hyperfine structure and isotope shifts are examples of the continuing role of 'conventional' spectroscopy. Even simple atoms give rise to complex spectra when electrons are excited from inner valence shells, and their study has helped in the development of many-body theories as well as in the understanding of recombination and scattering processes. The development of microwave and radiofrequency spectroscopy has allowed small energy differences such as those arising from hyperfine structure and quantum electrodynamical effects to be measured far more accurately than is possible by optical spectroscopy. The advent of tunable lasers has opened up a multitude of new techniques and, indeed, new fields of research concerned with the interactions of atoms with very high electromagnetic fields and non-linear optics. On the theoretical side the availability of powerful computers has enabled far more sophisticated calculations to be made. A high degree of co-operation between theory and experiment is vital in the interpretation of all these current fields of research.

The third application of spectroscopy, here designated spectrophysics, is concerned with interactions between atoms, molecules, electrons and photons insofar as these determine the intensities, widths and shifts of spectral lines and the intensity and spectral distribution of continuous radiation. There are two aspects to investigating these interactions: to interpret and understand what is happening, and to use this understanding to determine abundances, temperatures, pressures, velocities, electron densities and radiative transfer and dynamic rate processes in the emitting or absorbing gas. Spectrophysics has been an important branch of astrophysics ever since the identification of the Fraunhofer lines with applications both to true astrophysics and to what is sometimes called laboratory astrophysics. The importance of quantitative spectroscopy in both of these fields has increased rapidly in recent years for two main reasons. Until

1930 astrophysicists were limited in the range of wavelengths they could study by the 'optical window' in the Earth's atmosphere: 300 nm to 2 μm, in round figures. The opening of the 'radio window' (roughly 1 cm to 10 m range) set off the rapidly growing and immensely important science of radio astronomy. Of more direct relevance to the subject of this book, the development of rocket and satellite spectroscopy from 1945 onwards enabled 'optical' spectroscopists to see beyond the edges of their window to the infra-red and the highly important ultra-violet region below 300 nm. The second main reason for spectrophysical advance has been the development of high-temperature plasmas, initially for fusion work. The need for spectroscopic investigation of these plasmas stimulated lines of research that were immediately applicable to the study of astrophysical plasmas as well as to basic collisional and radiative procesess in atomic physics.

1.2 Scope of this book

This book is mainly concerned with those aspects of experimental spectroscopy that have definite physical applications – the third of the three uses described in the previous section. Chapters 2 and 3 survey atomic and molecular structure in order to give the necessary background for the rest of the book and to introduce the quantum numbers and the terminology. Chapters 4–7 describe the experimental techniques of optical spectroscopy, by which is meant spectroscopy from the soft X-ray region to the sub-millimetre region – the wavelength range from 1 nm to 100 μm (in round numbers). The emphasis in these chapters is on high resolution and intensity measurements, because these are necessary for the determinations of line shapes, transition probabilities and number densities. Absolute wavelength measurements and line identifications are of much greater importance in spectrochemistry and atomic structure than in spectrophysics. Chapter 8 surveys the current state of the continuously developing field of laser spectroscopy. Concluding the experimental techniques, Chapter 9 gives an overview of microwave, radiofrequency, level-crossing and Raman spectroscopy, with the object of pointing out alternative ways of measuring some of the quantities discussed in optical spectroscopy.

Chapter 10 discusses the broadening of spectral lines by radiation damping, Doppler effect and interactions with other particles. Chapter 11 defines and relates transition probabilities, absorption coefficients, oscillator strengths and connected quantities, and discusses such astrophysically important concepts as optical depth and curves of growth. Chapter 12 describes the various methods of measuring transition probabilities, partly because of their intrinsic importance in spectrophysics and partly because many such methods may equally well be used on lines of known transition probability to determine abundances and temperatures. Chapter 13 deals with energy distributions and equilibrium conditions and describes specifically some of the more important applications of spectroscopy to laboratory and astrophysical plasmas.

The book makes no attempt to cover a number of forms of spectroscopy – notably radio-astronomy, photoelectron, X-ray and Mössbauer spectroscopy – on the grounds either that they use very different experimental techniques or that they are not very relevant to spectrophysics.

1.3 The electromagnetic spectrum

The electromagnetic spectrum extends from the longest radio waves, whose wavelengths are measured in kilometres, through the medium and short radio waves to the microwave, infra-red and visible regions. Beyond the visible, the ultra-violet merges into the X-ray region, and at shorter wavelengths still come γ-rays and cosmic rays.

Figure 1.1 is a diagram of the spectrum on which are marked the conventional spectral regions. The divisions are necessarily rather arbitrary, except for the closely defined visible region. The wavelength λ is marked in metres and sub-units: micrometres, μm (or microns) and nanometres, nm. Ångström units, Å, and X-units, XU, are also shown since they are still recognized units and are frequently found in the literature. Since $1\ \text{Å} = 10^{-10}$ m, the nanometre and the ångström are related by $1\ \text{nm} = 10\ \text{Å}$. The XU was originally defined independently in such a way that its relation to the ångström unit was $1002.06\ \text{XU} = 1\ \text{Å}$. It is now defined as exactly 10^{-3} Å, or 10^{-13} m.

Despite the common classification by wavelength, the important quantity from the point of view of atomic and molecular structure is the frequency ν, which is related directly to the energy difference between two stationary states of the system by $\Delta E = h\nu$, where h is Planck's constant. Frequency is measured in Hz (s^{-1}) and multiples thereof (kHz, MHz, GHz, . . .) and is related to the wavelength measured in vacuum by

$$\nu = c/\lambda_{\text{vac}} \tag{1.1}$$

where c is the velocity of light in vacuum. In optical spectroscopy ν is usually replaced by a unit proportional to it, the wavenumber σ defined by

$$\sigma = \nu/c = 1/\lambda_{\text{vac}} \tag{1.2}$$

Different authors use different symbols for wavenumber: ν', $\tilde{\nu}$, k and σ will all be found in the literature, but we shall use σ in this book. The unit of wavenumber in the SI system should logically be the m^{-1}, but the unit cm^{-1} is firmly entrenched in actual usage. Attempts to designate it by the name Kayser (K) have not met with general acceptance, although the sub-unit millikayser (mK) is sometimes used in hyperfine structure and linewidth work. A convenient relation to remember is $1\ \text{mK} = 30\ \text{MHz}$. Finally, very high photon energies, corresponding to very short wavelengths, are often expressed in electron volts (eV): 1 eV is equal to 1.6×10^{-19} J and corresponds to 8066 cm^{-1}. These three 'energy units' are marked below the wavelength scale in Fig. 1.1, and the numerical relations are collected in the Appendix.

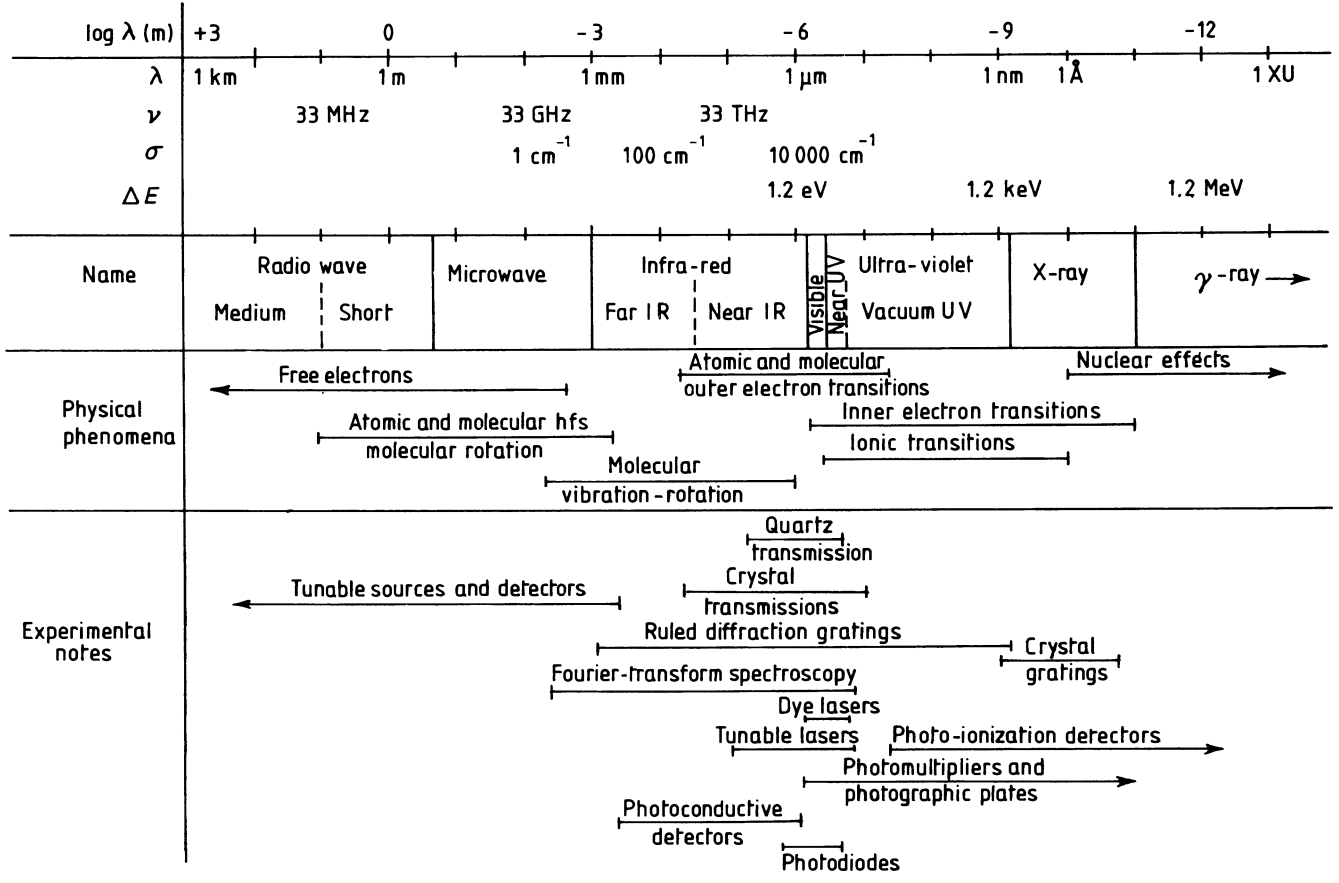

Figure 1.1 The electromagnetic spectrum.

The atomic and molecular energy levels themselves are also often expressed in wavenumbers, cm^{-1}. Conversion from absolute energy units, joules, requires in addition to the factor hc a further factor of 100 to change m^{-1} to cm^{-1}:

$$\Delta E\,(\text{joules}) = h\nu = hc/\lambda_{\text{vac}}\,(\text{m}) = 100\,hc\,\sigma(\text{cm}^{-1}) \tag{1.3}$$

The use of λ_{vac} should be stressed. If the wavelength is measured in air, it must be multiplied by n, the refractive index of air, before conversion to frequency or wavenumber units:

$$\nu = c/n\lambda_{\text{air}}, \qquad \sigma = 1/n\lambda_{\text{air}} \tag{1.4}$$

The correction is about 3 parts in 10^4, and is itself slightly wavelength-dependent. The tables for conversion of air wavelengths compiled by the National Bureau of Standards [21] include the refractive index correction in taking the reciprocal.

Returning to Fig. 1.1, it is seen that the visible region (750–400 nm) is bounded by the infra-red to longer wavelengths and the ultra-violet to shorter wavelengths. It is convenient to subdivide these regions into near and far infra-red at about 40 μm, and near and far ultra-violet at about 200 nm, these being roughly the practical limits of transparency of optical materials suitable for prisms and lenses. Until the middle of this century there were wavelength gaps, both between the far infra-red and radio and between the far ultra-violet and X-ray regions, in which it had not been found possible to excite and detect radiation. The invention of radar in the Second World War opened the door to the very short radio wave, or microwave, techniques, while at the same time infra-red spectroscopists were pushing up to longer wavelengths. The two now overlap in the sub-millimetre region. The short-wavelength gap has meanwhile been well filled, partly because of its interest to plasma- and astrophysicists. Optical spectroscopy now reaches wavelengths below 0.2 nm, while soft X-rays extend to 5 nm. The distinction between 'optical' and 'X-ray' in the overlap region is in the origin of the spectra. Optical spectroscopy is concerned with transitions of the outer or valence electrons, and X-ray spectroscopy with transitions of the innermost electrons. However, the distinction is somewhat blurred by 'inner shell' spectroscopy, dealing with excitation of electrons more tightly bound than the valence electrons. Very short wavelength 'optical' spectra may come either from these 'inner shell' electrons or from the outer electrons of very highly ionized species.

A book on optical spectroscopy may therefore take as a logical lower limit a wavelength just below 1 nm (a few ångströms). The upper limit is conventionally set in the far infra-red. However, the very smallest energy level splittings of interest are of order 0.01 cm^{-1}, corresponding to a wavelength of 1 m or so; this is the justification for including a chapter on microwave and radiofrequency spectroscopy. Apart from this one chapter, we shall be concerned with the conventional 'optical' region below 100 μm, with the emphasis on the three decades from 10 μm to 10 nm.

1.4 Term diagrams and line series

A spectrum consists of radiative transitions between stationary states of an atom or molecule. In the conventional energy level diagram these states are plotted vertically upwards in order of increasing energy. Figure 1.2 shows the simplest of all such diagrams, that of the hydrogen atom with the fine structure omitted. In this case the stationary states are defined by a single quantum number n. In optical spectroscopy the numerical value of the energy in units of cm^{-1} is known as the term value T.

Transitions between two stationary states i and f can be shown on a term diagram by vertical arrows, pointing up for absorption and down for emission. The wavenumber of the transition is simply the difference of the two term values: $\sigma = T_i - T_f$. Atomic spectroscopists are in the habit of writing transitions with the lower state first (regardless of whether they are looking at emission or absorption lines) while molecular spectroscopists, confusingly, write the upper state first.

The zero of energy is arbitrary because only energy differences can be measured. Conventionally, it is the state in which the electron and its parent ion

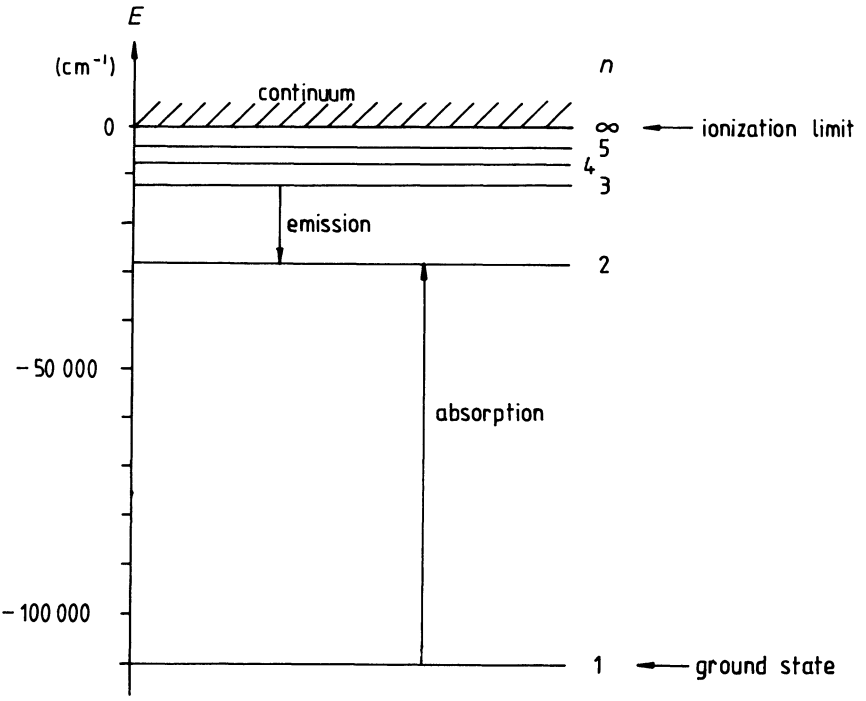

Figure 1.2 Energy level diagram of the hydrogen atom (with fine structure omitted). The energy levels are given by $E = -R_H/n^2$.

are at rest at infinite separation. The bound states are then states of negative energy. For the hydrogen atom they are given by the Bohr relation $T = -R_H/n^2$, where R_H is the Rydberg constant for hydrogen (109 678 cm^{-1}) and n is an integer – see section 2.1. The state of lowest energy (the ground state) has $n = 1$. As $n \to \infty$, $T \to 0$, and the terms crowd together as they approach the ionization limit, corresponding to the complete removal of the electron. The energy difference between this limit and the ground state, $100R_H hc$ joules, is the ionization energy of the hydrogen atom.

For states of positive energy, that is for the system ion + electron + kinetic energy, the energy is no longer quantized, and there exists a continuum of possible states, as shown for hydrogen in Fig. 1.2. Transitions between the continuum and a bound state and transitions within the continuum are possible and are known as free–bound and free–free transitions, respectively.

The hydrogen term diagram can equally well be taken to represent either the energy of the electron or that of the atom as a whole. In a many-electron atom it is the atom as a whole that matters. One can construct a term diagram by starting with all the electrons in their lowest states – the ground state of the atom – and then assuming one electron, normally the most weakly bound, to be excited to successively higher energies. Because of the interactions between the electrons, the other electrons are affected by this process, and the resulting stationary states are characteristic of the whole atom, not just of the excited electron. The resulting term diagram, though considerably more complex than Fig. 1.2, retains the essential features of one or more series of energy levels converging to ionization limits, the lowest of which corresponds to complete removal of the most loosely bound electron, leaving all other electrons in their ground states. If one of the more tightly bound electrons is excited, or two electrons at once, then an additional set of stationary states converges to a higher ionization limit corresponding to an excited state of the ion. This is illustrated schematically in Fig. 1.3.

The standard source of data for atomic and ionic energy levels is the National Bureau of Standards publication 'Atomic energy levels' and its various supplements [1]. These list all measured levels of every element for the neutral atom and for several stages of ionization. All terms in these tables are referred to the ground state of the relevant atom or ion as zero, so as to avoid inaccuracies from inexact values of ionization energies. An NBS bibliography [3] covers more recent work. Energy-level diagrams like that of Fig. 1.2 for the lighter atoms and ions are given in [2]. The most recent data for diatomic molecules are to be found in the compilations by Huber and Herzberg [10] and by Rosen [11].

When atoms or molecules are excited, for example in an electrical discharge, many of the energy levels are populated, and emission lines appear as these decay to lower states. Without such excitation only the ground state and possibly a few other low-lying states are appreciably populated, and absorption can take place only from those few states. Absorption spectra therefore show fewer lines than emission spectra. In either case the transitions between any one

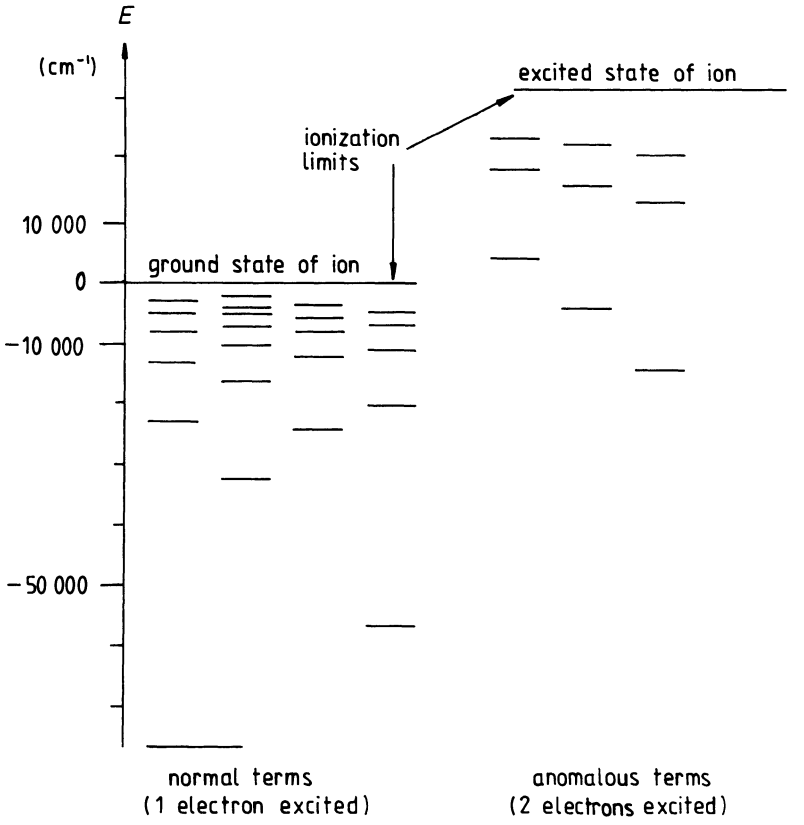

Figure 1.3 Schematic energy level diagram of a many-electron atom.

term and a succession of higher terms form a line series converging to some limit, which is identical with the ionization limit if the lower term is the ground state. Two such series are shown for hydrogen in Fig. 1.4. The Lyman series, for which the lower term has $n = 1$, converges to a limit in the far ultra-violet with $\sigma_\infty = R_H$, while the Balmer series, with $n = 2$ as the lower term, has its limit in the near ultra-violet ($\sigma_\infty = R_H/4$).

Beyond the series limit comes a region of continuous absorption or emission arising from the free–bound transitions. In absorption the process of removing an electron with finite kinetic energy is known as photo-ionization, while the inverse process in emission (recombination of an ion and an electron with emission of radiation) is known as radiative recombination.

There exist a number of compilations of wavelengths making it possible to identify the atom responsible for a line of measured wavelength. The most widely used are the *M.I.T. Wavelength Tables* [4], but other compilations cover

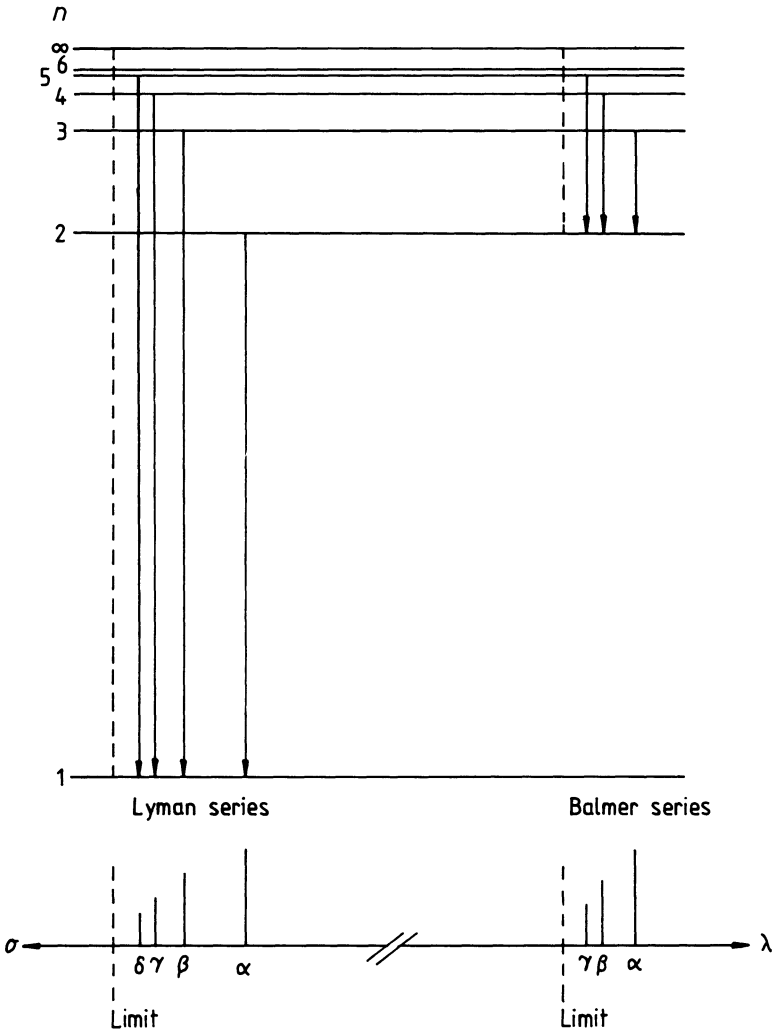

Figure 1.4 Line series in the spectrum of hydrogen.

particular classes of elements [5–7] and the infra-red [8] and vacuum ultra-violet [9] regions. Spectra of simple molecules may be identified from Pearse and Gaydon [12].

If energy levels (and the probability of transitions between them) are characteristic of atoms and molecules, how is it possible to deduce anything from an observed spectrum about the environment of the particle responsible for it? The observed intensity of a line depends on the distribution of population among the various excited states as well as on the intrinsic transition probabilities, and it is

also affected by absorption and re-emission by neighbouring atoms. Moreover, the energy levels themselves, and hence the spectral lines, may be broadened, shifted or split by interactions with external electric or magnetic fields or with the microfields due to neighbouring particles. Effects of this kind allow particle densities, temperatures, pressures and field strengths to be deduced from precise measurements of line shapes, shifts, widths and intensities.

1.5 Sources, detectors and dispersers

Figure 1.5 shows schematically the necessary elements of a spectroscopic system for (a) emission and (b) absorption. In emission the source is some form of arc, discharge or plasma in which the atom under study is excited, usually by electron collisions. In absorption the background source must have a continuous spectrum over the relevant wavelength region, and it is followed by an absorption cell. The disperser, or analyser, spreads the light out spatially according to its wavelength. Finally, the detector is the photographic plate, photoelectric detector or other device that records the intensity as a function of position. Only in Fourier transform spectroscopy is the information on wavelengths presented and analysed in an essentially different way.

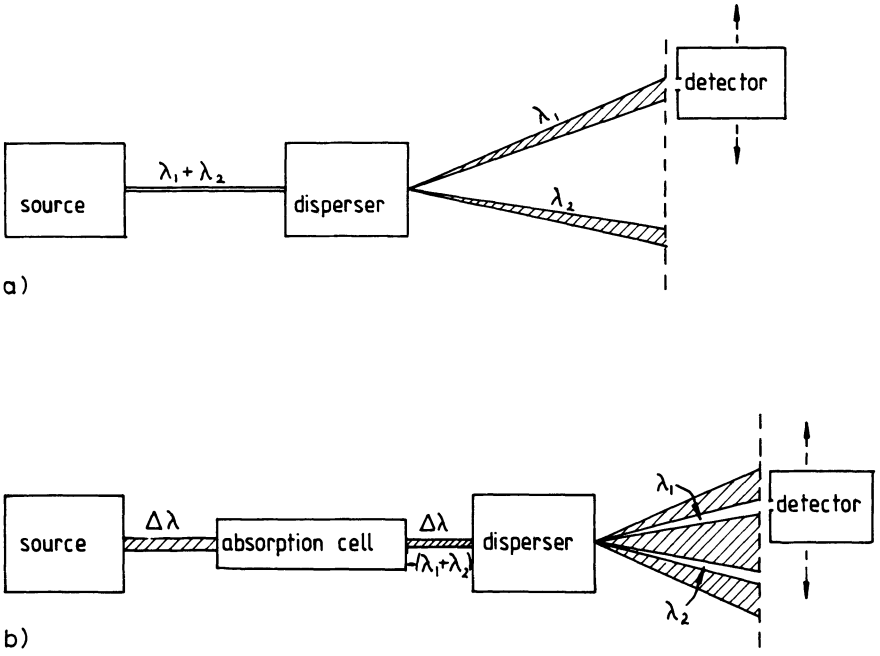

Figure 1.5 Schematic spectroscopic systems for (a) emission and (b) absorption.

In the microwave and radiofrequency regions the disperser is redundant because both the source and the detector can be tuned to any desired wavelength. The same is true in the optical region when the source is a tunable laser. However, most optical spectroscopy still makes use of dispersers, and it is worth making a few general points about these at this stage.

The dispersing element may be a prism, a grating or an interferometer; the corresponding instruments are treated in Chapters 5, 6 and 7, respectively. The only interferometers discussed in detail are the Fabry–Perot and the Michelson, the second of which is the instrument used in the important technique of Fourier transform spectroscopy. One can also classify spectroscopic instruments according to method of use as spectrographs, scanning spectrometers or monochromators. Etymologically, one should look through a spectroscope, take pictures with a spectrograph, measure (wavelengths or intensities) with a spectrometer and isolate light of a given colour with a monochromator. However, the term 'spectroscope', far from being confined to the visible region, is often used as a generic name, measurements are made with spectrographs as well as with spectrometers, and various imaging devices are now beginning to replace photographic plates for the picture-taking.

Figure 1.6 shows the essential features of a prism or a grating instrument. Light from the slit S is collimated by the lens L_1 and dispersed by the prism or grating G. The camera lens L_2 forms a set of images of S in its focal plane P in light of different wavelengths, and this constitutes the spectrum. A narrow slit is essential if lines that are close in wavelength are to be separated. In a spectrograph a photographic plate or imaging detector placed at P records many wavelengths simultaneously. A scanning spectrometer requires an exit slit S′ placed somewhere along P with an appropriate detector (photodiode, Golay cell, etc.) behind it; the spectrum is scanned either by tracking S′ and the detector along P, or by rotating the prism or grating. A very accurate drive is needed for

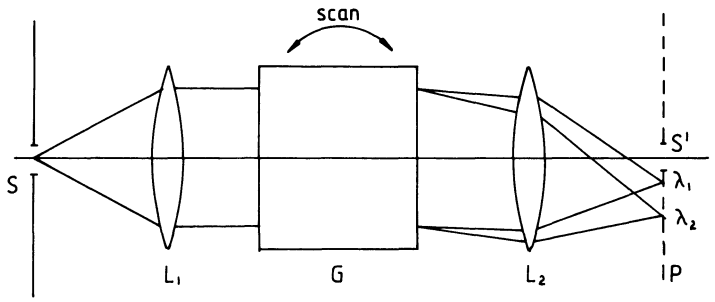

Figure 1.6 Essential features of prism or grating spectrometer. The collimating and 'camera' lenses L_1 and L_2 form a set of images S′ of the slit S in light of wavelengths λ_1, λ_2, . . . in the focal plane P of L_2. The spectral region is in general selected by rotating the prism or grating G.

high resolution work. If the instrument is used as a monochromator, with a relatively wide bandpass, the requirements on drive can normally be relaxed, but it is then convenient to design the instrument so that the exit slit is fixed and the emerging beam comes out at the same angle regardless of wavelength. Figure 1.6 is, of course, highly schematic, and it will be seen in Chapters 5 and 6 that a wide variety of arrangements is possible.

Figure 1.7 shows similarly the essential elements of interferometric spectroscopy. The lens L_1 collimates the light from the extended source S, and L_2 focuses the interference fringes from the interferometer I in the plane P. The important point of distinction is that the 'spectrum' is now a superposition of the interference patterns produced by each wavelength in the source, not a juxtaposition of geometric slit images. The interference pattern has circular symmetry, and the widths of the fringes are determined by the properties of the interferometer. When the Fabry–Perot is used photographically, the whole pattern, or some slice across it, is recorded on the plate at P. For the Fabry–Perot in scanning mode and for the Michelson, a circular aperture A isolates the central portion of the interference pattern, and the intensity is recorded while the optical path between the interferometer plates is changed.

Sources, detectors and 'dispersers' are interdependent: a disperser of high resolution may not be usable in practice if the source is weak and the detector insensitive. Conversely, the high resolution may be wasted if the lines are intrinsically very broad or are photographed on a very grainy plate. This dichotomy of intensity and resolution is always present in spectroscopy, and it is important to think about the balance in any particular experiment. It is considered more quantitatively in the following chapters.

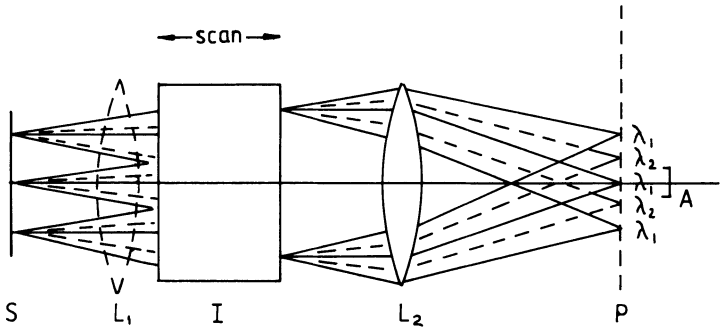

Figure 1.7 Essential features of interferometric spectrometer. L_1, L_2 and P are as in Fig. 1.6, but S is now an extended source, and the intensity in P is a superposition of interference patterns from the wavelengths λ_1, λ_2, . . . a portion of which is selected by the aperture A. In general the signal is recorded as a function of the spacing of the interferometer plates I.

1.6 Survey of spectral regions

We give here a brief summary of the experimental techniques that can be used in the different spectral regions, starting (illogically but chronologically) with the visible region and proceeding outwards. If we take 'visible' to reach a little way into both the infra-red and the ultra-violet, we have a well-defined region from 1000 nm (1 μm) to 200 nm, which is by far the easiest in which to work. Quartz is transparent throughout this region and glass is through most of it, so the choice is open between prisms, gratings and interferometers, and there is no difficulty about windows or lenses. The spectrum can be recorded photographically or photoelectrically. A wide range of sources is available for both emission and absorption spectroscopy, and among these one can now include tunable lasers (admittedly with certain constraints over the shortest wavelengths).

Going below 200 nm in wavelength, first air (or oxygen, to be precise) and then quartz start to absorb. To overcome the first the light path has to be evacuated – hence the name vacuum ultra-violet for this region. Quartz can be replaced by other crystals to extend the transmission range a few tens of nm, as far as 104 nm, which is the lithium fluoride limit, but prism spectrometers are used only for a few rather special applications in this region, and interferometers run into additional problems with surface tolerances and low reflectivity. Grating spectrometers are the instruments of real importance for wavelengths below 180 nm, with interferometers playing a marginal role down to about 120 nm. Lenses have to be replaced by mirrors or eliminated altogether by using concave gratings. Reflectivities decrease with wavelength, and below about 35 nm gratings have to be used at grazing incidence to overcome this problem. However, the high energy of the photons makes detection relatively easy; photographic or photoelectric methods may be used, and background problems are small compared with those of the infra-red. The crystal transmission limit at 104 nm often defines a jump in experimental complexity because of the additional difficulties of running light sources and absorption cells without windows. The regions above and below this barrier are often called after their first explorers – Schumann and Lyman, respectively. As for lasers, the incentive to extend the techniques of tunable laser spectroscopy into the vacuum ultra-violet region is such that any limitations stated here are certain to be out of date very soon; the state of the art at the time of writing is described in Chapter 8.

Going now the other way, into the infra-red, the effective limit of crystal transmission is about 40 μm, but prisms are little used above a few μm, and mirrors are so efficient that they are nearly always preferred to lenses. Grating and Fourier transform spectroscopy are the important techniques, and the interferometers for the latter can be made with thin films or grids to overcome the constraints of crystal beamsplitters. The principal problem of the infra-red is the low intensity of infra-red sources, together with background and noise problems that become increasingly severe as the photon energy approaches kT. It is usually necessary to trade spectral resolution off against signal-to-noise

ratio. However, a wide selection of lasers, both fixed frequency and tunable, is available and is increasing steadily. Finally, the vacuum requirements in the infra-red are not severe because dry oxygen and nitrogen do not absorb and it is only necessary to eliminate water vapour and carbon dioxide.

The infra-red overlaps with microwaves in the sub-millimetre region. Laser absorption spectroscopy simply becomes maser absorption spectroscopy, but in 'conventional' spectroscopy there is an abrupt change of technique to tunable oscillator sources and selective detectors without the need for dispersion. At still longer wavelengths, in the radiofrequency (rf) region, the detection of 'absorption lines' by the energy absorbed becomes increasingly insensitive. Other methods have been developed to detect the resonances, notably beam deflection and changes in optical radiation induced by the rf field, and this last category brings us back to the visible region.

1.7 Wavelength standards

To measure energy differences in a spectrum one needs to know the absolute wavelengths or frequencies of the transitions. Lasers have brought with them the possibility of making direct frequency measurements in the infra-red and visible regions at some hundreds of THz. Nevertheless, nearly all actual measurements are still derived from wavelength standards, and we discuss these first. A primary standard of wavelength is needed, together with a number of secondary standards distributed through the spectrum. A set of tertiary standards, readily reproducible in any laboratory, can then be generated from the secondary standards, and for virtually all purposes these are adequate for the wavelength calibration required by all prism and grating instruments. (Interferometers require fewer standards.) Some knowledge of these laboratory standards is useful, even if one is primarily concerned with measuring line shapes and widths rather than absolute wavelengths, and the saga of the rise and fall of the primary wavelength standard is in itself of sufficient interest to warrant a brief account.

Until 1983 the primary standard of wavelength was an orange line of krypton emitted under certain specified conditions, and most of the existing secondary standards have been measured interferometrically against this standard in different laboratories with accuracies of order $1:10^7$. For the most part they are lines of iron, thorium and the inert gases, which are updated and added to every few years by a commission of the International Astronomical Union [14]. A much larger number of tertiary standards has been derived from these, often with similar interferometric accuracy [15–20]. However, accuracy degenerates as one goes into the infra-red or ultra-violet, because of difficulties associated with the large number of intermediate steps and with the actual measurements. The lack of good traditional standards in the infra-red is for some purposes compensated by the availability of laser lines, but in the vacuum ultra-violet it is particularly unfortunate because so many series limits and important resonance

lines are found in this region. Overlapping orders in grating spectra have been used to extend visible standards into the ultra-violet, but this method is liable to large systematic errors. Much more reliable are 'Ritz standards', wavelengths calculated from the Ritz combination principle: the frequency ν_{13} corresponding to a transition 1–3 is obtained from the sum of the frequencies to an intermediate level 2, $\nu_{12}+\nu_{23}$, where the transitions 1–2 and 2–3 fall in the visible region [19,20].

The history of the primary wavelength standard is inseparably tied up with the history of the metre. Until 1960 the ultimate standard of length was the standard metre, the distance between two scratches on a platinum–iridium bar kept in Paris. There were obvious drawbacks to such a standard. The first was geographical and became evident as early as 1884, when it was shown that the great chart of absolute wavelengths compiled by Ångström some 16 years earlier, based on the Uppsala metre, needed correcting by 13 parts in 10^5, the discrepancy between the two metre bars. The second disadvantage became apparent with the development of high-resolution spectroscopy at the end of the nineteenth century: the accuracy of wavelength measurement was actually limited by the finite width of the scratches on the bar. To remedy this it was decided to adopt as an international standard of *wavelength* the red line of cadmium, emitted from a source in certain specified conditions. Once this had been determined as accurately as possible in terms of the Paris metre, all other wavelengths could be measured relative to it. However, the standardization of the red cadmium line revealed the third defect of the Paris metre: measurements over a period of years showed the bar to be shrinking slowly as the result of metal fatigue. (Similar measurements on the Imperial Yard in London revealed an even worse shrinkage.) All this led to the decision to adopt a wavelength as the ultimate standard of *length*. The red cadmium line was no longer the best choice, because line broadening arising from unresolved isotope shift in lines more suitable in other respects could be eliminated by using separated isotopes. After some years of controversy over the relative merits of the green line of mercury and the orange line of krypton, the latter was chosen in 1960, and the international standard of length was defined to 4 parts in 10^9 by taking 1 m to be 1 650 763.73 vacuum wavelengths of this line ($\lambda_{vac} = 605.780\,210$ nm) emitted by the isotope ^{86}Kr under certain specified conditions.

As a matter of historical interest, the mercury green line which was the runner-up in the contest was one of the lines investigated by Michelson as a possible standard in the 1890s, using visibility measurements with his own interferometer. He rejected it in favour of the red cadmium line because of what we now recognize as isotope structure.

The 1960 decision is not the end of the saga. Since 1967 the fundamental standard of time, or frequency, has also been an atomic standard, defined by a transition between the hyperfine structure levels of the ground state of ^{133}Cs. The second is defined as 9 192 631 770 periods of this transition frequency, and the measurement can be made to $1:10^{13}$, which corresponds to 1 second in

300 000 years [13]. This 'atomic clock' is described briefly in Chapter 9. The measurement and comparison of frequencies is the most accurate of all types of measurement, and with the development of lasers it has become possible to transfer this frequency standard down a multistage chain of oscillators and lasers into the near infra-red [13]. At the end of the chain come helium–neon lasers, stabilized by locking on to various molecular transitions in methane and iodine, notably those at 3.39 μm, 1.15 μm and 633 nm (Chapter 8). With frequencies in the visible known to about $1:10^{10}$ and the ability to compare laser wavelengths interferometrically to a few parts in 10^{10}, the velocity of light can be measured to an accuracy limited only by the krypton-generated metre. Since only two out of the three quantities, λ, ν and c can be separately defined, the limitation on wavelength measurements imposed by the krypton standard can be removed by abolishing the metre altogether and fixing the value of c. This has now officially happened; in 1983 c was *defined* as 299 792 458 $\mathrm{m\,s^{-1}}$, and the metre has become the distance travelled by light in a fraction $1/c$ of an 'atomic second'. The wavelength of any line is the reciprocal of its frequency, multiplied by the constant c. Any line that is a frequency standard automatically becomes also a wavelength standard, capable of comparison with another wavelength to a few parts in 10^{10}, an improvement of an order of magnitude over the krypton standard.

With the important exception of Doppler-free laser spectroscopy, described in Chapter 8, the ultimate limits of accuracy are purely of academic interest in spectroscopy, as distinct from metrology. For 'thermal' light sources the best interferometric measurements, made to a fraction of the Doppler-broadened line width, are of order $1:10^{7}$.

1.8 Units used in this book

All expressions and equations are in the form appropriate to SI units, but I have not stuck rigorously to these in the text when other units are more convenient or more widely used in practice. A few relations are given in CGS units as well as SI, and where possible I have kept $(e^2/4\pi\varepsilon_0)$ bracketed as a unit so that anyone working in CGS units can simply replace it by e^2. In any case the conversion can easily be made by replacing ε_0, wherever it appears, by $1/4\pi$.

I have abandoned the ångström unit with considerable reluctance, in view of its convenience as a unit of atomic size as well as its familiarity as a wavelength label. Beware the factor of 10 between Å and nm when comparing with older and some current literature; astrophysicists, in particular, tend still to use ångströms.

Energy differences present an anomaly. In general, practising spectroscopists express energy differences in $\mathrm{cm^{-1}}$, although MHz and GHz are frequently used for small differences and eV for very large ones. As remarked in section 1.3, the SI unit of wavenumber should be the $\mathrm{m^{-1}}$, but *nobody* actually uses it. The

consequence, as in equation 1.3, is the appearance of a factor of 100 in the conversion of cm^{-1} to any other energy unit.

In the literature of plasma physics and astrophysics number densities are still given in cm^{-3} rather than in m^{-3}. I have often quoted number densities in both units, not because I believe the reader to be incapable of multiplying by 10^6, but because it seems useful to get a feeling for the orders of magnitude met in practice. For the same reason I have occasionally quoted values of magnetic induction in gauss as well as in tesla.

Finally, I should make clear my distinction between approximations, represented by the symbol $\simeq$, and orders of magnitude, represented by $\sim$. Useful physical insight can frequently be gained from order-of-magnitude arguments when more exact calculations are too difficult, too lengthy or actually impossible, and the symbol $\sim$ is usually the printed equivalent of a verbal 'arm-waving argument'.

References

ATOMIC ENERGY LEVELS

1. Moore, C.E. (1949–58) Atomic energy levels, *NBS Circ. 467*, Vols I, II and III. (Reprinted 1971 as *NSRDS–NBS 35.*)

These tables have been supplemented by

(i) Moore, C.E. (1965–85) Selected tables of atomic spectra, *NSRDS–NBS 3*, published as separate sections, each covering one or more stages of ionization of one element and containing as Part A revised energy levels and as Part B multiplet tables; sections 1–11.
(ii) Martin, W.C., Zalubas, R. and Hagan, L. (1978) Atomic energy levels–the rare earth elements, *NSRDS–NBS 60.*
(iii) Sugar, J. and Corliss, C. (1985) Atomic energy levels of the iron-period elements: potassium through nickel, *J. Phys. Chem. Ref. Data*, **14** (suppl. 2).
(iv) Various authors, articles published in *J. Phys. Chem. Ref. Data* from **2** (1973) onwards, each giving energy levels for many stages of ionization of one element.
(v) Moore, C.E. (1970) Ionization potentials and ionization limits derived from the analyses of atomic spectra, *NSRDS–NBS 34.*

Additional sources of information are

2. Bashkin, S. and Stoner, J.O. (1975) *Atomic Energy Levels and Grotrian diagrams*, Vol. 1, H I–P XV, North-Holland, Amsterdam.

and the comprehensive NBS bibliography

3. Hagan, L. and Martin, W.C. (1968–83) Bibliography on atomic energy levels and spectra, *NBS Spec. Publ. 363* (original and three supplements).

ATOMIC WAVELENGTHS

4. Harrison, G.R. (1959) *M.I.T. Wavelength Tables*, Wiley, New York.
5. Moore, C.E. (1959) Multiplets of astrophysical interest, *NBS Tech. Note 36*; (1950–62) Ultra-violet multiplet tables, *NBS Circ. 488.*
6. Meggers, W.F., Corliss, C.H. and Scribner, B.F. (1975) Tables of spectral line intensities: part I–arranged by elements; part II–arranged by wavelength, *NBS Monogr. 145.*

7. Striganov, A.R. and Sventitskii, N.S. (1968) *Tables of Spectral Lines of Neutral and Ionized Atoms*, IFI/Plenum, New York. (22 elements of interest in plasma spectroscopy.)
8. Outred, M. (1978) Tables of atomic spectral lines for the 10 000 Å to 40 000 Å region, *J. Phys. Chem. Ref. Data*, **7**, 1.
9. Kelly, R. L. (1973) Atomic and ionic emission lines below 2000 Å: H–Kr, *Natn. Res. Lab. Rep. 7599*; (1982) Atomic and ionic spectrum lines below 2000 Å: H–Ar, *Oak Ridge Natn. Lab. Rep. 5922*; (1979) Atomic emission lines in the near ultra-violet: H–Kr, *NASA Tech. Mem. 80268.*

CONSTANTS OF DIATOMIC MOLECULES

10. Herzberg, G. (1950) *Spectra of diatomic molecules*, Van Nostrand, New York. The data in the appendix of this book has been expanded, updated and published separately as: Huber, K.P. and Herzberg, G. (1979) *Constants of Diatomic Molecules*, Van Nostrand, New York.
11. Rosen, B. (1970) *Spectroscopic Data Relative to Diatomic Molecules* (*International Tables of Selected Constants*, Vol. 17), Pergamon Press, New York.
12. Pearse, R.W.B. and Gaydon, A.G. (1963) *The Identification of Molecular Spectra*, Chapman and Hall, London.

WAVELENGTH STANDARDS

For a discussion of primary standards, see

13. Petley, B. W. (1985) *The Fundamental Physical Constants and the Frontiers of Measurement*, Hilger, London.

Secondary standards are published from time to time as a report of Commission 14 of the IAU.

14. *Trans. Int. Astr. Un.*

Lists are also to be found in

15. *Handbook of Chemistry and Physics*, CRC Press, Cleveland, Ohio (published annually).

More than 5000 lines are listed in

16. Kaufman, V. and Edlen, B. (1974) Reference wavelengths from atomic spectra in the range 15 Å–25 000 Å, *J. Phys. Chem. Ref. Data*, **3**, 825.

Additional sources of infra-red standards are

17. Cole, A.R.H. (1977) Tables of wavenumbers for the calibration of infrared spectrometers, *IUPAC Commission on Molecular Structure and Spectroscopy*, Pergamon Press, Oxford.
18. Humphreys, C.J. (1974) First spectra of neon, argon and xenon in the 1.2–4.1 μm region, *J. Phys. Chem. Ref. Data*, **2**, 519.

Additional sources of ultra-violet standards are

19. Samson, J.A.R. (1967) *Techniques of Vacuum Ultra-violet Spectroscopy*, Wiley, New York.
20. Zaidel, A.N. and Schreider, E.Y.A. (1970) *Vacuum Ultra-violet Spectroscopy*, Humphrey, Ann Arbor, Michigan.

WAVELENGTH CONVERSION (air wavelengths to vacuum wavenumbers)

21. Coleman, C.D., Bozmann, W.R. and Meggers, W.F. (1960) Table of wavenumbers, *NBS Monogr. 3.*

TWO

A summary of atomic structure

Before discussing the methods of spectroscopy, it is useful to summarize the main features of atomic and molecular spectra in order to see what effects one is likely to be investigating in any given spectral region. Some understanding of structure and of the terminology used to describe it is also a necessary prerequisite to the later chapters on spectrophysics. A two-chapter survey, though necessarily superficial, will at least define some of the terms used in later chapters. Selections of books covering atomic and molecular structure in proper detail, at different degrees of rigour, are listed at the end of this and the next chapter.

2.1 Schrödinger's equation: wave functions

It is well known that the energy levels of the hydrogen atom, neglecting fine structure, are given by the simple Bohr formula $T = -R_H/n^2$, where R_H is a constant (the Rydberg constant) and n is a positive integer. Bohr derived this relation by assuming the electron to follow a circular orbit round the nucleus, obeying the laws of classical mechanics with two exceptions: the orbital angular momentum must be an integral multiple of $\hbar(=h/2\pi)$, and in such an allowed orbit the electron does not radiate. It is easily shown that for a two-body system – electron plus nucleus – the energy of a state having angular momentum $n\hbar$ is

$$E_n = -\frac{\mu e^4}{2(4\pi\varepsilon_0)^2\hbar^2}\frac{Z^2}{n^2} \quad \text{(joules)} \tag{2.1}$$

where μ is the reduced mass of the electron, $mM/(m+M)$, M is the mass of the nucleus and Z is its charge (1 for H, 2 for He^+, etc.). If the constants are in SI units, dividing E (joules) by $100ch$ to convert to wavenumbers gives for the

Rydberg constant

$$R=\frac{2\pi^2\mu e^4}{100ch^3(4\pi\varepsilon_0)^2}\quad \text{cm}^{-1} \tag{2.2}$$

which is numerically equal to 109 678 cm^{-1} for hydrogen. For an infinitely heavy nucleus $\mu\to m$, and $R_\infty = 109\,737.31\ \text{cm}^{-1}$. The dependence of μ on M gives rise to isotope shift (section 2.11), but for most purposes the distinction between μ and m can be ignored.

The radii of the Bohr orbits are given by

$$a_n=\frac{(4\pi\varepsilon_0)\hbar^2}{me^2}\frac{n^2}{Z}\equiv\frac{n^2}{Z}a_0 \tag{2.3}$$

where a_0, the radius of the first Bohr orbit, is treated as the 'atomic unit' of length. It is numerically equal to 5.29×10^{-11} m, or 0.529 Å.

Rather than discuss the modifications and additions to Bohr's theory, we shall move straight on to an outline of quantum-mechanical treatment. Of the two possible formulations, wave mechanical and matrix, the first will be used here.

The starting point is Schrödinger's equation

$$-\frac{\hbar^2}{2m}\nabla^2\psi+V\psi=i\hbar\frac{\partial\psi}{\partial t} \tag{2.4}$$

Here ∇^2 is the Laplacian operator, V is the potential energy of the electron and ψ is the wave function. $\psi(r, t)$ is to be interpreted as a complex 'probability amplitude' such that $\psi^*\psi$, or $|\psi|^2$, is the probability of finding the electron at the point in space defined by the co-ordinates r, θ, φ (Fig. 2.1) at time t. The nucleus is the origin of co-ordinates.

The operator on the left-hand side of the equation, $-(\hbar^2/2m)\nabla^2+V$, is the quantum-mechanical operator **H**, which can be derived from the classical Hamiltonian, $(p^2/2m)+V$, by substituting for the components of linear momentum:

$$p_x\to-i\hbar\frac{\partial}{\partial x},\quad \text{etc.}$$

For a stationary state of an isolated atom V is independent of time; both sides of the equation can then be set equal to a constant E, and ψ can be factored as

$$\psi(r, t)=u(r)e^{-iE/\hbar t} \tag{2.5}$$

The spatial part of the wave function u is the solution of Schrödinger's time-independent equation

$$-\frac{\hbar^2}{2m}\nabla^2u+(E-V)u=0 \tag{2.6}$$

which expresses the conservation of total energy E.

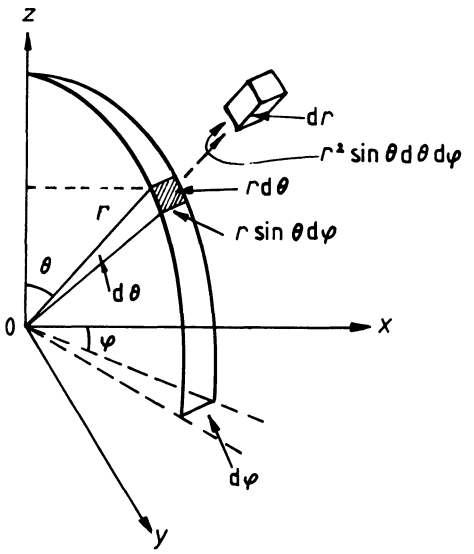

Figure 2.1 Spherical polar co-ordinates.

Equation 2.6 written in the form

$$\mathbf{H}u = Eu \tag{2.7}$$

is known as an eigenvalue equation. A function u_i satisfying this equation is an eigenfunction of the Hamiltonian with a corresponding eigenvalue E_i. Evidently the us are undetermined to the extent of an arbitrary multiplying constant: this is chosen to normalize the function, so that

$$\int u_i^* u_i \, \mathrm{d}\boldsymbol{r} = 1 \tag{2.8}$$

where $\mathrm{d}\boldsymbol{r}$ represents the volume element $\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$ or $r^2 \sin\theta\,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}\varphi$ (see Fig. 2.1) and the integral is taken over all space. The eigenfunctions have the property of orthogonality, such that

$$\int u_j^* u_i \, \mathrm{d}\boldsymbol{r} = 0 \quad \text{for } i \neq j \tag{2.9}$$

Finally, it follows from their interpretation as probability densities that the average, or expectation value, of any observable a (position, momentum, etc.) is found from the corresponding operator $\mathbf{A}$ by

$$\langle a \rangle \equiv \bar{a} = \int u_i^* \mathbf{A} u_i \, \mathrm{d}\boldsymbol{r} \tag{2.10}$$

For example, the average value of r for a given state represented by u_i is given by $\langle r \rangle = \int r u_i^* u_i \, \mathrm{d}\boldsymbol{r}$ since the operator $\mathbf{r}$ is identical with r. As a widely used notation, the integral in equation 2.10 is written in the form $\langle i|\mathbf{A}|i\rangle$.

2.2 The hydrogen atom: degeneracy and spin

For hydrogen and, more generally, for any one-electron ion of nuclear charge Z, V is the Coulomb potential

$$V = -\frac{Ze}{(4\pi\varepsilon_0)r} \tag{2.11}$$

and Schrödinger's equation, like its classical counterpart, has an exact analytical solution. With ∇^2 expressed in spherical polar co-ordinates and u factorized into radial and angular parts,

$$u(r, \theta, \varphi) = R(r)\, Y(\theta, \varphi) \tag{2.12}$$

equation 2.6 splits into two equations, one involving functions of r only and the other functions of θ and φ only. Certain restrictions are imposed on the mathematically possible solutions in order to make u behave in a physically sensible fashion: it must be single-valued, continuous and finite everywhere. It is these restrictions that introduce the quantum numbers. The solutions to the angular equation are the well-known spherical harmonics

$$Y_l^m = \text{const.}\, e^{im\varphi} P_l^m(\cos\theta) \tag{2.13}$$

where $P_l^m(\cos\theta)$ are the associated Legendre polynomials. The constraints are that m and l both be integers, with $m = 0, \pm 1, \pm 2, \ldots$ and $l \geqslant |m|$. The constant is chosen to satisfy the normalization condition

$$\int_0^{2\pi} \int_0^{\pi} Y_l^{m*}\, Y_l^m \sin\theta \, d\theta \, d\varphi = 1 \tag{2.14}$$

The integer l also appears in the radial equation, the solution to which is another polynomial. The condition that $R(r) \to 0$ as $r \to \infty$ (i.e. that the polynomials have a finite number of terms) is met for negative E by the condition

$$E_n = -\frac{1}{2} \frac{e^4 m}{(4\pi\varepsilon_0)^2 \hbar^2} \frac{Z^2}{n^2} \tag{2.1a}$$

where n is an integer satisfying

$$n \geqslant l + 1$$

For positive E there is no restriction. Equations 2.1 and 2.1a are identical for $\mu \to m$.

The functions $R_{n,l}(r)$ depend on the two integers n and l and satisfy the normalization condition

$$\int_0^{\infty} R_{n_l}^* R_{n_l} r^2 \, dr = 1 \tag{2.15}$$

Tables 2.1 and 2.2 give, respectively, the functions Y_l^m and $R_{n,l}$ for the first few values of n, l and m, and Figs 2.2 and 2.3 show their behaviour graphically.

Table 2.1 Angular hydrogen wave functions (spherical harmonics) Y_e^m

$$Y_0^0 = \sqrt{\left(\frac{1}{4\pi}\right)}$$

$$Y_1^0 = \sqrt{\left(\frac{3}{4\pi}\right)}\cos\theta \quad Y_1^{\pm 1} = \mp\sqrt{\left(\frac{3}{8\pi}\right)}\sin\theta e^{\pm i\varphi}$$

$$Y_2^0 = \sqrt{\left(\frac{5}{16\pi}\right)}(3\cos^2\theta - 1) \quad Y_2^{\pm 1} = \mp\sqrt{\left(\frac{15}{8\pi}\right)}\sin\theta\cos\theta e^{\pm i\varphi}$$

$$Y_2^{\pm 2} = \sqrt{\left(\frac{15}{32\pi}\right)}\sin^2\theta e^{\pm 2i\varphi}$$

normalized to make $2\pi\int_{\theta=0}^{\pi} \sin\theta\, d\theta Y^* Y = 1$

Table 2.2 Radial hydrogen wave functions $R_{n,l}$

(1s) $$R_{1,0} = \left(\frac{Z}{a_0}\right)^{3/2} 2e^{-Zr/a_0}$$

(2s) $$R_{2,0} = \left(\frac{Z}{2a_0}\right)^{3/2} 2\left(1 - \frac{Zr}{2a_0}\right)e^{-Zr/2a_0}$$

(2p) $$R_{2,1} = \left(\frac{Z}{2a_0}\right)^{3/2} \frac{2}{\sqrt{3}}\left(\frac{Zr}{2a_0}\right)e^{-Zr/2a_0}$$

(3s) $$R_{3,0} = \left(\frac{Z}{3a_0}\right)^{3/2} 2\left[1 - 2\frac{Zr}{3a_0} + \frac{2}{3}\left(\frac{Zr}{3a_0}\right)^2\right]e^{-Zr/3a_0}$$

(3p) $$R_{3,1} = \left(\frac{Z}{3a_0}\right)^{3/2} \frac{4\sqrt{2}}{3}\left(\frac{Zr}{3a_0}\right)\left(1 - \frac{1}{2}\frac{Zr}{3a_0}\right)e^{-Zr/3a_0}$$

(3d) $$R_{3,2} = \left(\frac{Z}{3a_0}\right)^{3/2} \frac{2\sqrt{2}}{3\sqrt{5}}\left(\frac{Zr}{3a_0}\right)^2 e^{-Zr/3a_0}$$

normalized to make $\int_0^{\infty} r^2 dr\, R^* R = 1$

$$\left(a_0 = \text{1st Bohr radius} = \frac{4\pi\varepsilon_0\hbar^2}{me^2} = 5.29 \times 10^{-11}\,\text{m}\right)$$

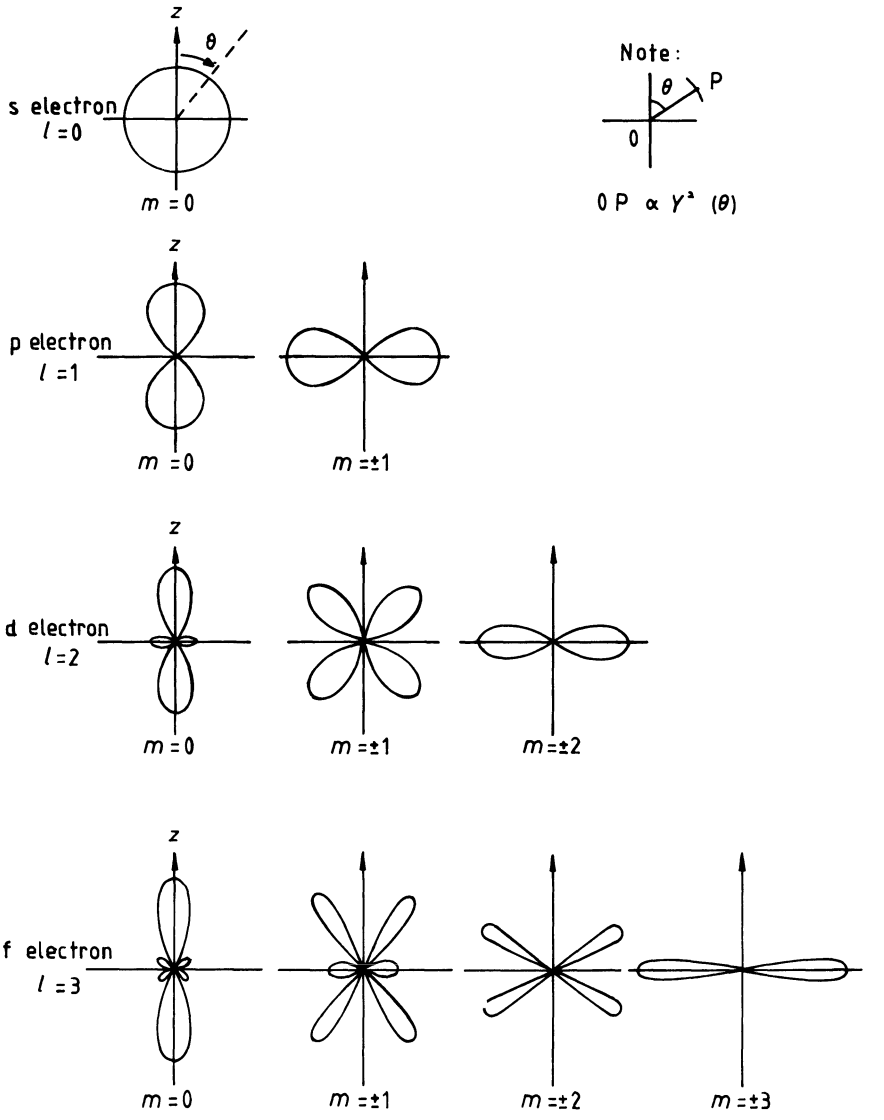

Figure 2.2 The angular functions Y^*Y for $l=0$–3. The length of the line joining the origin to the curve at a particular value of θ is proportional to Y^2 for that value of θ. All the functions are symmetric about the z-axis.

The 'allowed' values for the energy given by equation 2.1 are the eigenvalues of equation 2.7. The allowed values of orbital angular momentum, found from the corresponding quantum mechanical operators, are $\sqrt{[l(l+1)]}\hbar$ for the total orbital angular momentum and $m\hbar$ for its component in a fixed direction in

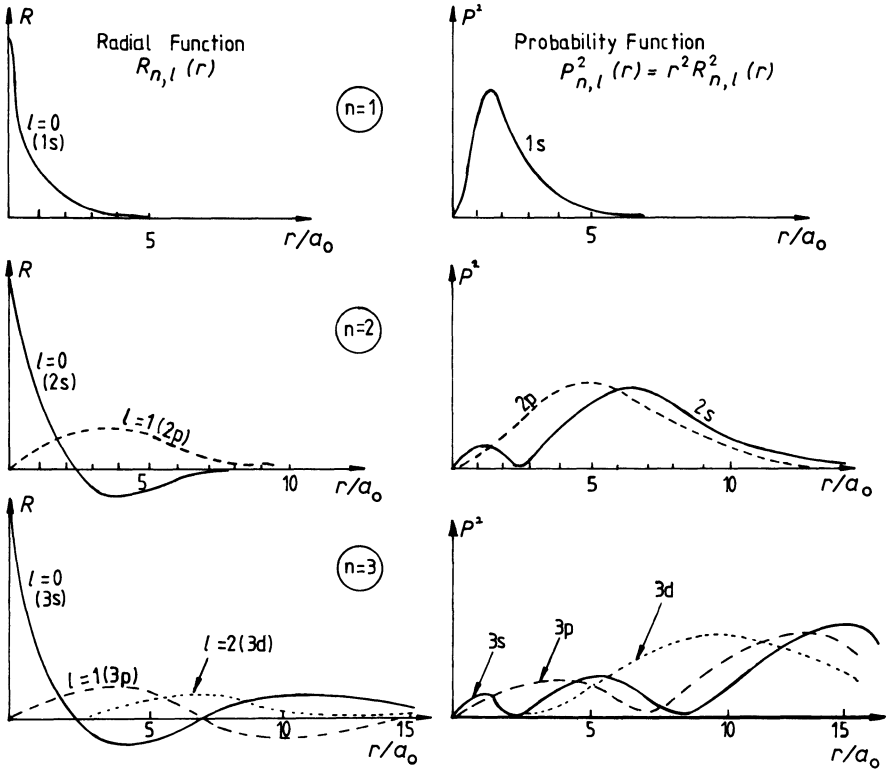

Figure 2.3 Hydrogen radial wave functions. The actual functions $R(r)$ are on the left, and the curves on the right show the probability of finding the electron at a distance r from the nucleus.

space (conventionally the z-axis). From the constraints on l and m given above, it follows that for a given n (and hence a given energy) there are n values of orbital angular momentum $-l=0, 1, \ldots, (n-1)$ – and for each value of l there are $2l+1$ values of the z-component $-m=l, l-1, \ldots, 0, \ldots, -l$. The orbital angular momentum vector is a source of possible confusion: it is often written as $l\hbar$, but its actual magnitude is $\sqrt{[l(l+1)]}\hbar$, often designated by l^*h. For large l, obviously $l^* \to l$. As the maximum component of $l\hbar$ along the z-axis is lh, not $l^*\hbar$, l can never point exactly in the z-direction. This is illustrated in Fig. 2.4 for $l=2$.

The wave function u depends on all three of the quantum numbers n, l and m; for fixed n, each combination of l and m describes a different electron distribution. It can be seen from the φ-dependence of the spherical harmonics that u^*u is always independent of φ, so the electron distribution always has cylindrical symmetry about the z-axis. For $l=0$, Y_l^m, and hence Y^*Y, is also independent of θ, so in this particular case the electron distribution is spherically symmetric. The other important characteristic of the wave function for $l=0$ is that $R(r)$ does

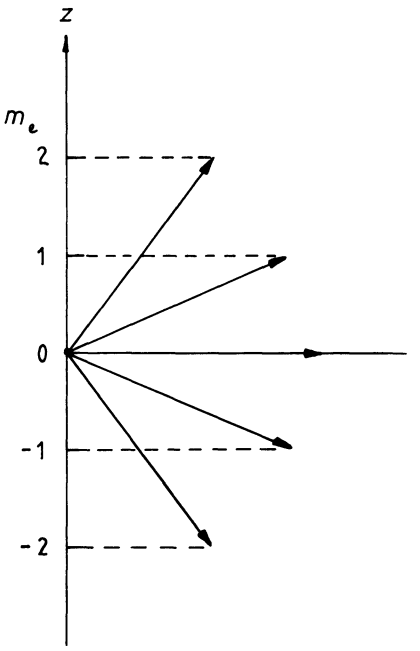

Figure 2.4 Relation between the quantum numbers l and m_l for $l=2$. The length of the l-vector, l^*, is $\sqrt{6}$, or 2.45.

not vanish at the origin, and there is therefore a finite probability of finding the electron near the nucleus.

Different wave functions having the same energy are said to be degenerate. Since E depends on n only, the degeneracy of the state n is

$$\sum_{l=0}^{n-1}(2l+1)=n^2$$

The m-degeneracy is easily understood, for in the absence of an external field there is no reason for singling out any one direction in space. The l-degeneracy is a direct result of the Coulomb potential and is peculiar to hydrogen. When relativistic effects and spin–orbit interaction are included, it is actually partly removed, as described in section 2.7.

A fourth quantum number is required to specify the state of the electron completely. Certain effects in atomic spectra – notably, fine structure and the Zeeman effect – can be explained by attributing to the electron an intrinsic spin angular momentum $s\hbar$, of magnitude $\sqrt{[s(s+1)]}\hbar$ (or $s^*\hbar$), where s always has the value $\frac{1}{2}$. The component of spin in the z-direction, m_s, is also quantized: m_s can take only the values $+\frac{1}{2}$ ('spin up') or $-\frac{1}{2}$ ('spin down'). m_s is the fourth

quantum number. In the absence of an external field the energy does not depend on m_s, so (still ignoring fine structure) the energy levels of the hydrogen atom are $2n^2$-fold degenerate. Electron spin is a natural consequence of the full relativistic treatment of the electron first carried through by Dirac. It can be included formally in the Schrödinger wave function by writing the latter as $\Phi(r, \theta, \varphi, s)$. When spin–orbit interactions are neglected this can be rewritten as the product of the spatial function $u(r, \theta, \varphi)$ and a spin function $\sigma(s)$,

$$\Phi(r, \theta, \varphi, s) \rightarrow u(r, \theta, \varphi)\, \sigma(s)$$

2.3 The Pauli exclusion principle and shell structure

The many-body problem (nucleus plus two or more electrons) is not exactly soluble in either classical or quantum mechanics. As a first approximation one assumes each electron to move independently in some (non-Coulomb) potential and to have a wave function that can be specified by the same four quantum numbers as in the hydrogen atom. The basis for this assumption is discussed in the next section.

Were it not for the Pauli exclusion principle, all the electrons of every atom in the ground state would have $n=1$, corresponding to minimum energy, and the table of the elements would not be a periodic table. Electrons are fermions, and the exclusion principle forbids any two of them to occupy the same state – i.e. to have all four of their quantum numbers identical. This limits the number of electrons that can have the same principal quantum number n. For each of the n possible values of l there are $2l+1$ different values of m (from now onwards this quantum number will be designated m_l), and for each of these two possible values of m_s, totalling $2n^2$, as has already been seen. Table 2.3 shows the vacancies available for the first three values of n. Each cell in the penultimate line may contain two, and only two, electrons. The nomenclature in the bottom line is important: electrons having $l=0$, 1, 2, 3 are known as s, p, d, f electrons, respectively. 2p, for example, means that the electron has $n=2$, $l=1$. This terminology is historical rather than logical; it comes from the labels attached to the first line series identified in the spectra of the alkali metals: sharp, principal, diffuse and fundamental. After f the order becomes alphabetical. The notation is unfortunate in that the s has nothing whatever to do with the spin angular momentum vector $\boldsymbol{s}$.

The *electron configuration* describes the values of n and l allocated to all the electrons. The number of electrons in each sub-shell is shown as a superscript. For example, the configuration for eight electrons in their lowest energy states is $1s^2 2s^2 2p^4$. If one electron is excited to 3s, the configuration becomes $1s^2 2s^2 2p^3 3s$.

In the Bohr theory of the hydrogen atom the radius of the electron orbit increases as n^2. In wave mechanics the expectation value for r derived from the

Table 2.3

n	1	2					3									
l	0	0	1				0	1			2					
m_l	0	0	+1	0	−1		0	+1	0	−1	+2	+1	0	−1	−2	
m_s	↑↓	↑↓	↑↓	↑↓	↑↓		↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	
Name	1s	2s	2p				3s	3p			3d					

electron probability distribution still depends to a first approximation on n^2. All electrons with the same n may therefore be regarded as forming a sort of 'shell' around the nucleus. This concept has even greater validity when applied to the 'sub-shells' formed by electrons having identical values of n *and* l. There are $2(2l+1)$ electrons in each sub-shell, as seen from the table. A full sub-shell has several important features. Firstly, the pairing of positive and negative values of m_l and m_s means that there is no resultant angular momentum. Secondly, the wave functions are such that the resultant electron distribution is spherically symmetric. (This follows from summing $Y_l^{m*} Y_l^m$ in equation 2.13 over all m_l for a given l.) Thirdly, the structure is a very stable one; all of the electrons in the sub-shell have the same radial charge distribution (the function $R(r)$ depends only on n and l, and is independent of m_l), and consequently the electrons screen each other only partially from the nuclear charge. As one goes along a row of the periodic table, adding one electron and one proton per element, the effective positive charge seen by the electrons increases as the sub-shell fills, and their binding energy increases correspondingly. This is shown experimentally by the steady rise in ionization potential along the row up to the inert gas that corresponds to the filling of the last vacancy.

Closed shells and sub-shells may therefore be thought of as forming an inert 'core' surrounding the nucleus, and the behaviour of an atom is determined mainly by the electrons in unfilled sub-shells, the valence electrons. The core screens most of the nuclear charge from the valence electrons, so long as the latter do not have an appreciable probability of being found within it. It is this spherically symmetric core that justifies the central field approximation.

2.4 The central field model

The exact form of Schrödinger's equation for an atom with more than one electron includes in the potential energy V not only the attractive interaction between each electron and the nucleus, $-Ze^2/(4\pi\varepsilon_0)r_i$, but also the repulsive interactions, $+Ze^2/(4\pi\varepsilon_0)r_{ij}$, between each pair of electrons. These cross-terms prevent separation into equations each involving co-ordinates of only one

electron, which can be solved exactly. The problem is tackled by assuming each electron to be subject to a potential V that represents the combined effect of the nucleus and all the other electrons. If, further, one takes a spherical average over the latter, so that V is a function of r only and not of angle, one arrives at the central field model. This has the enormous advantage that the radial and angular variables may be separated as in the case of hydrogen. Moreover, the θ and φ equations are identical with those of hydrogen, since V does not appear in them, and they have the same solutions Y_l^m. The R equation is different because $V(r)$ no longer has the simple Coulomb form. The solution still specifies the quantum number $n \geqslant l+1$, but the energy is determined by the exact form of $V(r)$ and now depends on l as well as n. The problem of calculating a potential $V_i(r)$ for the ith electron, which depends on the (unknown) radial wave functions of all the other electrons, is usually tackled by some sort of self-consistent field approach. One guesses a plausible wave function for each electron, computes the corresponding $V(r)$ from Poisson's equation, and solves Schrödinger's equation to obtain a new set of $R(r)$. The cycle is repeated until V and R become consistent. Even without going into these calculations one can understand why the energy depends on l as well as n, which is the most important distinction between the central field atom and the hydrogen atom. Regardless of the exact form of V, the behaviour of the radial wave functions is such that only an s-electron ($l=0$) has a finite probability of being found close to the nucleus. Figure 2.5 shows how the radial probability depends on l in a typical case for $n=3$, $l=0$, 1, 2. The probability distribution for the core electrons, $n=1$ and $n=2$, is also shown on the diagram. It is the s-electrons that have the greatest probability of penetrating the core; they therefore 'see' on average a higher positive charge, and so are more closely bound. Electrons with the maximum value of l, d-electrons in this case, are mostly well outside the core and have almost hydrogen-like energies and wave functions. Figure 2.6 shows a typical form of energy level diagram for a single valence electron. The different l-values are conventionally displaced sideways in order of increasing l.

The spherically averaged potential is a good approximation for closed subshells, and the central field model obviously works best when there is only one valence electron, as in the alkali metals. Atoms with two or more valence electrons are found to have several different possible energy states, or terms, for any given electron configuration. To account for these one must look more specifically at the direct electrostatic repulsion between the valence electrons. There is also another interaction that has so far been ignored, in hydrogen as well as in the central field model: the magnetic interaction between the orbital angular momentum and the intrinsic electron spin angular momentum. In considering these effects one usually starts with the electrostatic interaction, because in many cases of practical interest it is the larger of the two. We may justify the order in which these three interactions – central field, residual electrostatic and magnetic – are taken by quoting values typical of one of the first excited states of a simple atom: central field energy $\simeq 5$ eV, residual electrostatic

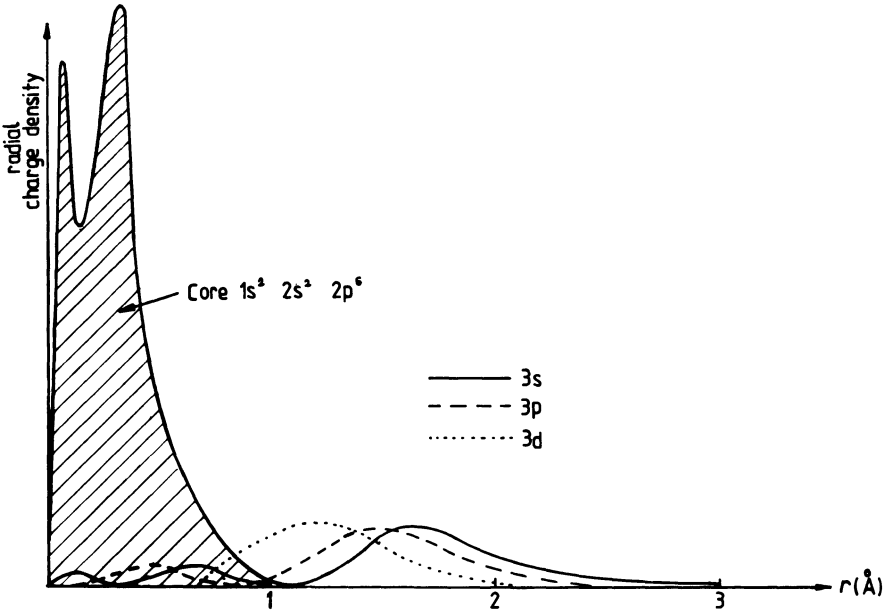

Figure 2.5 Radial charge densities of the $n=1$ and $n=2$ core electrons and of the $n=3$ valence electron for different values of l. The 3d-electron is almost entirely outside the core.

energy $\simeq$ 1–2 eV, magnetic energy from $\sim 10^{-3}$ eV in light atoms to $\sim 10^{-1}$ eV in medium atoms (often comparable with electrostatic energy in heavy atoms).

2.5 Interlude: perturbation theory

The central field model is the first stage of approximation to the many-electron problem. How does one apply corrections, to this or any other approximation, to arrive at a more realistic solution? In answering this question for the central field, we are concerned only with stationary states of the atom, and therefore with time-independent corrections, or 'perturbations'; but as transitions between stationary states involve time-dependent perturbations both types will have to be considered.

We start with the time-independent case and the corresponding Schrödinger equation in the form of equation 2.7:

$$\mathbf{H}\varphi = E\varphi \tag{2.16}$$

The operator $\mathbf{H}$ is the sum of the zero-order operator $\mathbf{H}^0$ and the perturbation operator $\mathbf{H}^1$ representing the correction term, and φ and E are its eigenfunctions

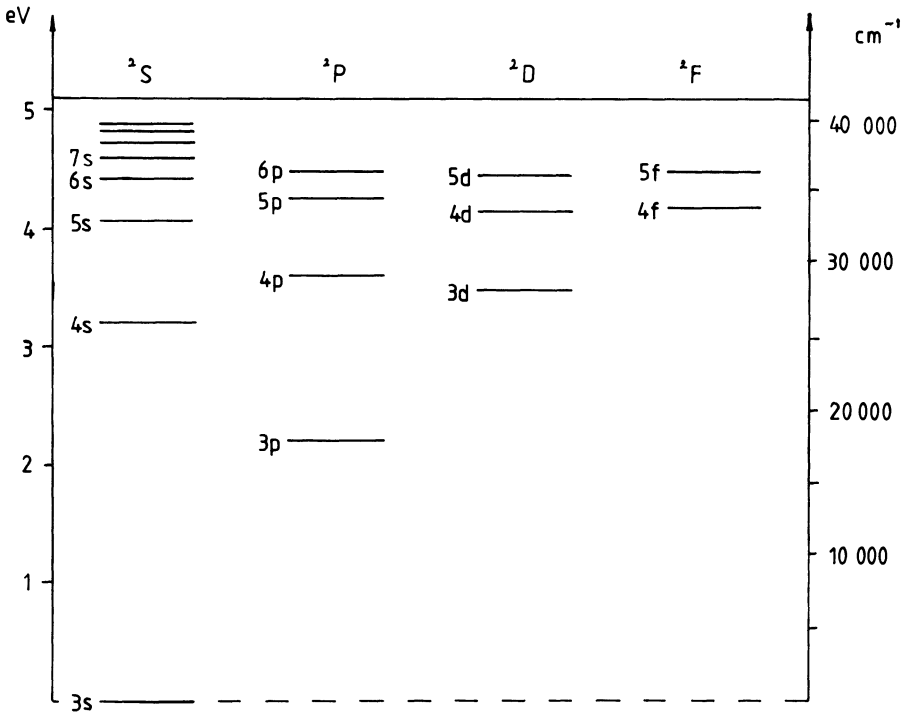

Figure 2.6 Dependence of energy levels on l. The energy level diagram is that of sodium with the fine structure omitted.

and eigenvalues, both unknown. Equation 2.16 can be written as

$$(\mathbf{H}^0+\mathbf{H}^1)\varphi=(E^0+E^1)\varphi \tag{2.17}$$

where E^0 is the zero-order eigenvalue and E^1 is the correction we wish to determine. The essence of perturbation theory is that if $E^1 \ll E^0$ the functions φ are not very different from the eigenfunctions u_i satisfying $\mathbf{H}^0 u_i = E^0 u_i$. It can then be shown (see any text on quantum mechanics) that E^1 is the expectation value of the perturbation $\mathbf{H}^1$ evaluated with the *unperturbed* wave functions, i.e.

$$E^1=\int u_i^* \mathbf{H}^1 u_1 \, d\boldsymbol{r} \equiv \langle i|\mathbf{H}^1|i\rangle \tag{2.18}$$

The theory can be carried to second and higher orders by evaluating the corrections to the eigenfunctions and using these new functions to find the second-order energy correction, and so on. The new functions are formed by mixing in to a given zero-order function u_i a proportion of all other zero-order functions:

$$u_i^1 = u_i + \sum_{k \neq i} c_k u_k \tag{2.19}$$

The coefficients c_k are given by

$$c_k = \frac{\langle k|\mathbf{H}^1|i\rangle}{E_i - E_k} \tag{2.20}$$

where E_k is the unperturbed energy of the kth state. The second iteration, evaluating $\mathbf{H}^1$ with the functions of equation 2.19 leads to a correction to the energy:

$$E^{(2)} = \sum_{k \neq i} c_k \langle i|\mathbf{H}^1|k\rangle = \sum_{k \neq i} \frac{\langle k|\mathbf{H}^1|i\rangle\langle i|\mathbf{H}^1|k\rangle}{E_i - E_k} \tag{2.21}$$

Time-*dependent* perturbation theory will be necessary for the treatment of transitions between stationary states. We start from equation 2.4 in the form

$$[\mathbf{H}^0 + \mathbf{H}^1(t)]\psi = \mathrm{i}\hbar \frac{\partial \psi}{\partial t} \tag{2.22}$$

where $\mathbf{H}^1(t)$ is the time-dependent perturbation. The consequent time variation of ψ is expressed by writing it as a linear combination of stationary-state eigenfunctions ψ_k of $\mathbf{H}^0$ with time-dependent coefficients so that, using equation 2.5,

$$\psi(\boldsymbol{r}, t) = \sum_k c_k(t)\,\psi_k(\boldsymbol{r}, t) = \sum_k c_k(t)\,u_k(\boldsymbol{r})\,\mathrm{e}^{-\mathrm{i}E_k/\hbar t}$$

Using this expression in equation 2.22, together with the eigenvalue equation $\mathbf{H}^0 u_k = E_k u_k$, which is valid for all k, we get

$$\mathrm{i}\hbar \sum_k \dot{c}_k\,u_k\,\mathrm{e}^{-\mathrm{i}E_k/\hbar t} = \sum_k c_k\,\mathbf{H}^1\,u_k\,\mathrm{e}^{-\mathrm{i}E_k/\hbar t}$$

The system is assumed to start in a particular state (i): $c_i(0) = 1$ and all other $c_k(0) = 0$. Then for *small* perturbations, such that all $c_k(t)$ remain small and c_i remains approximately equal to 1, we can ignore all the terms on the right-hand side except for the ith, $\mathbf{H}^1 u_i \exp(-\mathrm{i}E_i/\hbar t)$. The trick used to isolate a single c_k on the left-hand side is to multiply through by u_k^* and integrate over all space, taking advantage of the normalization and orthogonality relations of equations 2.8 and 2.9. Each $c_k(t)$ is then determined by a differential equation independent of all the others:

$$\mathrm{i}\hbar\,\dot{c}_k\,\mathrm{e}^{-\mathrm{i}E_k/\hbar t} \simeq \int u_k^*\,\mathbf{H}^1\,u_i\,\mathrm{e}^{-\mathrm{i}E_i/\hbar t}\,\mathrm{d}\boldsymbol{r}$$

To get any further, one needs the explicit time dependence of $\mathbf{H}^1(t)$. The equation is then integrated to give $c_k(t)$, and the probability of finding the system in state k after time t is

$$P_{ik}(t) = |c_k(t)|^2 \tag{2.23}$$

where

$$c_k(t) = -\frac{\mathrm{i}}{\hbar}\int_0^t \mathrm{d}t\,\mathrm{e}^{\mathrm{i}[(E_k - E_i)/\hbar]t}\,\langle k|\mathbf{H}^1|i\rangle$$

2.6 Electrostatic interactions: the quantum numbers L and S

The central field potential $V_i(r)$ in which any outer electron moves comprises not only the contributions of the nucleus and core electrons, but also the greater part of the potential due to other electrons in the same sub-shell. The residual electrostatic interaction caused by departures from spherical symmetry in the distribution of these other electrons has to be applied as a correction to the central field energy. If the correction is small, it can be calculated by first-order perturbation theory from the unperturbed central field wave functions, as described in section 2.5.

To get a qualitative picture, we will first consider the 'vector model' for two valence electrons. The magnitude of the electrostatic repulsion between them depends on an integral of the form

$$\int\int u_i^*(1)\, u_j^*(2) \frac{e^2}{(4\pi\varepsilon_0) r_{12}} u_i(1)\, u_j(2)\, \mathrm{d}\boldsymbol{r}_1\, \mathrm{d}\boldsymbol{r}_2$$

where $u_i(1)$ and $u_j(2)$ are the wave functions of the two electrons, and the integral is taken over all space. For a given electron configuration, $(n, l)_1$ and $(n, l)_2$ are fixed by definition, but the value of the integral is different for different values of $(m_l)_1$ and $(m_l)_2$. This can be understood physically from Fig. 2.2: the ms determine the planes in which the electron distributions are concentrated and hence affect the average value of $1/r_{12}$. The dependence of the interaction on the relative orientation of the $\boldsymbol{l}$-vectors can be described by defining a resultant orbital angular momentum vector $\boldsymbol{L}\hbar$, formed by the vector addition $\boldsymbol{L}=\boldsymbol{l}_1+\boldsymbol{l}_2$ (Fig. 2.7). L acts as an additional quantum number for the two-electron system, taking integral values l_1+l_2, $l_1+l_2-1, \ldots, |l_1-l_2|$. As with l, the actual magnitude of the total orbital angular momentum is $\sqrt{[L(L+1)]}\hbar$, or $L^*\hbar$. Its component in the z-direction, M_L, can take the $2L+1$ integral values from $+L$ to $-L$. States with different L-values are called terms, and are labelled with the capital letters S, P, D, . . . for $L=0, 1, 2, \ldots$ by analogy with the single-electron terminology. For example, two p-electrons can give rise to S, P and D terms having $L=0$, 1 and 2, respectively. If one of the electrons is an s-electron ($l_1=0$), then $L=l_2$, and there is no splitting into terms of different L. This is consistent with the spherical symmetry of the charge distribution of an s-electron.

In the case of three or more electrons the total orbital angular momentum is defined in the same way by $\boldsymbol{L}=\boldsymbol{l}_1+\boldsymbol{l}_2+\boldsymbol{l}_3+\ldots$, and L may in general take any integral value from the maximum $\Sigma|l|$ down to zero.

The splitting of a configuration into different terms in accordance with the relative orientation of the ls is essentially a classical effect, though evaluated for the wave-mechanical charge distributions. However, it is found that a given configuration splits into more terms than can be accounted for by the range of possible L-values. In particular, an atom with two valence electrons has two distinct sets of terms, called singlets and triplets from the way in which they

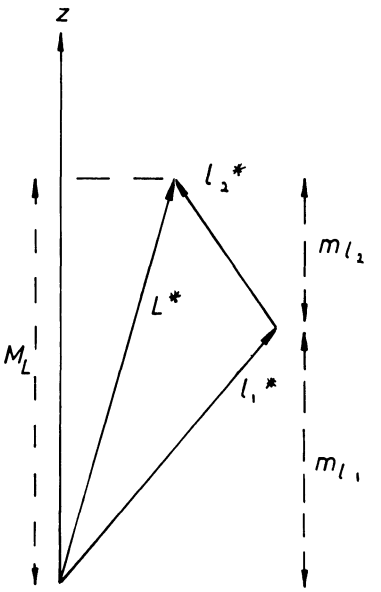

Figure 2.7 Vector model for the quantum number L. The lengths of the arrows represent the magnitudes of the vectors, with $l_1^* = \sqrt{[l_1(l_1+1)]}$, etc.

behave when the magnetic interactions are taken into account. These two sets are distinguished by an additional quantum number S, defined by $\boldsymbol{S} = \boldsymbol{s}_1 + \boldsymbol{s}_2$ and having the two possible values $S = 1$ and $S = 0$. $\boldsymbol{S}\hbar$ is the total spin angular momentum vector of the system, of magnitude $\sqrt{[S(S+1)]}\hbar$ (or $S^*\hbar$). The multiplicity of a term is given by $2S+1$, for reasons that are given in the next section, so $S = 1$ corresponds to the triplet terms and $S = 0$ to the singlet. The multiplicity is shown as a superscript to the left of the letter giving the L-value; for example, the full list of terms arising from the configuration of two p-electrons is ^{1}S, ^{1}P, ^{1}D, ^{3}S, ^{3}P, ^{3}D.

Since the two sets are distinguished by their S-values, it might seem reasonable to attribute the energy difference between them to spin–spin interaction. However, this is a magnetic dipole–dipole interaction which turns out to be too weak to account for the observed energy differences. The explanation is to be found in an exchange interaction that has no classical counterpart. Its effect is that the system can exist in two different states in which the motions of the electrons are differently correlated, so that the average *electrostatic* interaction between them is also different. Each of these states is associated with one of the values of S for reasons of symmetry. To see how this comes about, let us take the configuration *n*s,*n*p. This has only one possible value of L ($L = 1$), so we are concerned only with ^{1}P and ^{3}P terms. In the zero-order approximation, in which the non-central part of the interaction between the two electrons is

ignored, the wave function for the system can be written as the product of the two independent electron wave functions, $U = u_s(1)u_p(2)$. But this wave function is not physically satisfactory because if we interchange the electrons we get a different wave function $u_p(1)u_s(2)$, and hence a different electron distribution, for an identical system – identical because the electrons are indistinguishable. (Remember that (1) and (2) stand for the electron co-ordinates, and from inspection of Fig. 2.3 it is clear that the product of an s-function at r_1 with a p-function at r_2 is not in general equal to that of a p-function at r_1 with an s-function at r_2.) A physically acceptable wave function has to change either not at all or else only in sign, so that the actual electron distribution $|U|^2$ is invariant against electron interchange. The two linear combinations

$$\begin{aligned} U_+ &= \frac{1}{\sqrt{2}}[u_s(1)u_p(2) + u_p(1)u_s(2)] \\ U_- &= \frac{1}{\sqrt{2}}[u_s(1)u_p(2) - u_p(1)u_s(2)] \end{aligned} \tag{2.24}$$

satisfy this requirement: the first is symmetric and does not change, and the second is antisymmetric and changes sign. Both may be thought of as describing a situation in which both electrons are partly in an s-state and partly in a p-state. When one comes to evaluate the perturbation energy from equation 2.18 using either U_+ or U_-

$$E^1 = \iint U_\pm^* \frac{e^2}{(4\pi\varepsilon_0)r_{12}} U_\pm \, d\boldsymbol{r}_1 \, d\boldsymbol{r}_2$$

it can be seen that in addition to straightforward Coulomb integrals of the form

$$\iint u_s^*(1)\, u_p^*(2) \frac{e^2}{(4\pi\varepsilon_0)r_{12}} u_s(1)\, u_p(2)\, d\boldsymbol{r}_1 \, d\boldsymbol{r}_2 \Rightarrow \frac{1}{4\pi\varepsilon_0} \iint \frac{\rho(r_1)\,\rho(r_2)}{r_{12}} d\boldsymbol{r}_1 \, d\boldsymbol{r}_2$$

there must also be cross-terms of the form

$$\iint u_s^*(1)\, u_p^*(2) \frac{e^2}{(4\pi\varepsilon_0)r_{12}} u_p(1)\, u_s(2)\, d\boldsymbol{r}_1 \, d\boldsymbol{r}_2$$

which have no classical interpretation. These are known as exchange integrals. Setting the Coulomb integral equal to J and the exchange integral equal to K, one finds for the perturbation energies of the U_+ and U_- states, respectively:

$$\Delta E_+ = J + K \qquad \Delta E_- = J - K \tag{2.25}$$

The exchange interaction splits the configuration into two terms differing in energy by $2K$. Both J and K are repulsive, and therefore positive, terms, so the U_+ state lies above the U_- state (Fig. 2.8). One can interpret this as a correlation in the electron motions. It is obvious from inspection of U_- that the probability of finding the electrons close together in this state is vanishingly small since

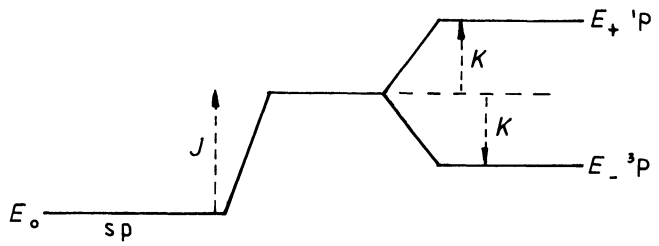

Figure 2.8 Electrostatic interaction in an sp-configuration. J is the (classical) Coulomb energy and K the (non-classical) exchange energy.

$U_- \to 0$ as co-ordinates (1)→co-ordinates (2). It can be proved explicitly that in the U_- state the electrons are on average further apart.

It remains to show the connection of the spin with all this. The Pauli principle can be restated in the form that the total wave function must be antisymmetric against electron exchange. To satisfy this, we must combine an antisymmetric spin function σ_- with U_+ and a symmetric spin function σ_+ with U_-. Table 2.4 shows schematically how the four possible spin combinations give rise to one antisymmetric and three symmetric spin functions. The pairs (b) and (c) of column 1 have to be recombined as in column 2 to form spin functions that are invariant or simply change sign when the electrons are interchanged. Of the four functions in this column, only (c) is antisymmetric with respect to electron interchange. The component of total spin in the z-direction, M_S of column 4, is the result of taking the upward-pointing arrows to define the direction of positive z. The three σ_+ functions can be attributed to the three possible orientations of a total spin $S=1$ and must be combined with U_-, forming a triplet term. The one σ_- function is associated with $S=0$ and is combined with U_+ to give the singlet term. In this approximation the ^3P term is triply degenerate, because in the absence of a field there is nothing to distinguish the three M_S values. The degeneracy is removed when the spin–orbit interaction is introduced.

Table 2.4

	1 Electron spins	2 Spin function	3 Symmetry	4 M_s	5 S	Degeneracy	Orbital function
a	↑↑	↑↑	σ_+	+1 ↘			
b	↑↓	↑↓+↓↑	σ_+	0 →	1	triplet	U_-
c	↓↑	↑↓−↓↑	σ_-	0 ↗			
d	↓↓	↓↓	σ_+	−1	0	singlet	U_+

In a configuration such as pp the Coulomb J-integral has different values for each of the possible L-values representing S, P and D terms. Each of these terms is then split by the appropriate K-integral into a singlet and a triplet term, giving different energies for each of ^{1}S, ^{3}S, ^{1}P, ^{3}P, ^{1}D and ^{3}D. For *equivalent electrons* – that is, electrons having the same values of both n and l – some terms are forbidden by the Pauli exclusion principle because they require both electrons to have the same values of m_l and m_s. In this instance only ^{3}P, ^{1}S and ^{1}D are allowed.

When there are three or more valence electrons, S is defined by $\boldsymbol{S}=\boldsymbol{s}_1+\boldsymbol{s}_2+\boldsymbol{s}_3+\ldots$, and takes integral or half-integral values according to whether the number of electrons is even or odd: singlets, triplets and quintets for four electrons, for example, and doublets and quartets for three electrons. As with two electrons, the terms that can be formed from equivalent electrons are limited by the Pauli exclusion principle.

2.7 Magnetic interactions: the quantum numbers j and J

The electron has an intrinsic magnetic moment $\boldsymbol{\mu}_s$ by virtue of its charge and spin angular momentum $\boldsymbol{s}\hbar$. The classical value of the gyromagnetic ratio would suggest the relation

$$\boldsymbol{\mu}_s=-\frac{e}{2m}\boldsymbol{s}^*\hbar$$

but the experimental value for $\boldsymbol{\mu}_s$, confirmed theoretically by Dirac's relativistic treatment, is almost exactly twice as large. This 'anomalous' gyromagnetic ratio is expressed in terms of the Bohr magneton μ_B and a 'g-factor' by

$$\boldsymbol{\mu}_s=-g_s\mu_B\boldsymbol{s}^* \tag{2.26}$$

where $\mu_B=e\hbar/2m=0.927\times10^{-23}\,\mathrm{J\,T^{-1}}$ (in wavenumber units, $\mu_B=0.467\,\mathrm{cm^{-1}\,T^{-1}}$). g_s can be taken as 2 for most purposes, although it is actually greater by about $1:10^3$.

This magnetic moment is the cause of the fine structure observed in atomic spectra. Consider first an atom with one valence electron. As a result of its motion in the central electric field, the electron experiences a magnetic field $\boldsymbol{B}_l$ proportional to its angular momentum $\boldsymbol{l}^*\hbar$. The resulting energy is given by

$$\Delta E_{ls}=-\mu_s\cdot\boldsymbol{B}_l$$

or

$$\Delta E_{ls}=a\boldsymbol{s}^*\cdot\boldsymbol{l}^* \tag{2.27}$$

where a is a constant that can be calculated if the central field potential is known. The angular momenta $\boldsymbol{l}\hbar$ and $\boldsymbol{s}\hbar$ are coupled by this interaction to form a resultant total angular momentum $\boldsymbol{j}\hbar$, where $\boldsymbol{j}=\boldsymbol{l}+\boldsymbol{s}$. Like the other angular momenta, this is quantized, with the quantum number j restricted to the values $l+\frac{1}{2}$ and $l-\frac{1}{2}$. The actual magnitude of the vector $\boldsymbol{j}\hbar$ is $\sqrt{[j(j+1)]}\hbar$, or $j^*\hbar$; thus

the two possible values of j correspond respectively to $\boldsymbol{l}$ and $\boldsymbol{s}$ not quite parallel and not quite antiparallel, as shown in Fig. 2.9(a). It can be seen from this figure that the scalar product $\boldsymbol{l}\cdot\boldsymbol{s}$ in equation 2.27 is given by the cosine rule:

$$\boldsymbol{l}^*\cdot\boldsymbol{s}^* = \tfrac{1}{2}(j^{*2} - l^{*2} - s^{*2})$$

The two possible values of E_{ls} corresponding to the two possible j-values give rise to the doublet structure in the spectra of the alkali metals and other atoms with one valence electron. The fine structure levels are labelled with the j-values as a subscript: $^2P_{1/2}$ and $^2P_{3/2}$ from a p-electron, $^2D_{3/2}$ and $^2D_{5/2}$ from a d-electron, etc. For an s-electron there is no splitting, and $^2S_{1/2}$ is the only level. The fine structure splitting is small compared with the difference between levels of different l, except in the hydrogen atom where the levels of different l are degenerate.

When $\boldsymbol{l}$ and $\boldsymbol{s}$ are coupled in this manner, the quantum numbers m_l and m_s cease to be meaningful; $\boldsymbol{l}$ and $\boldsymbol{s}$ can be regarded as precessing around their

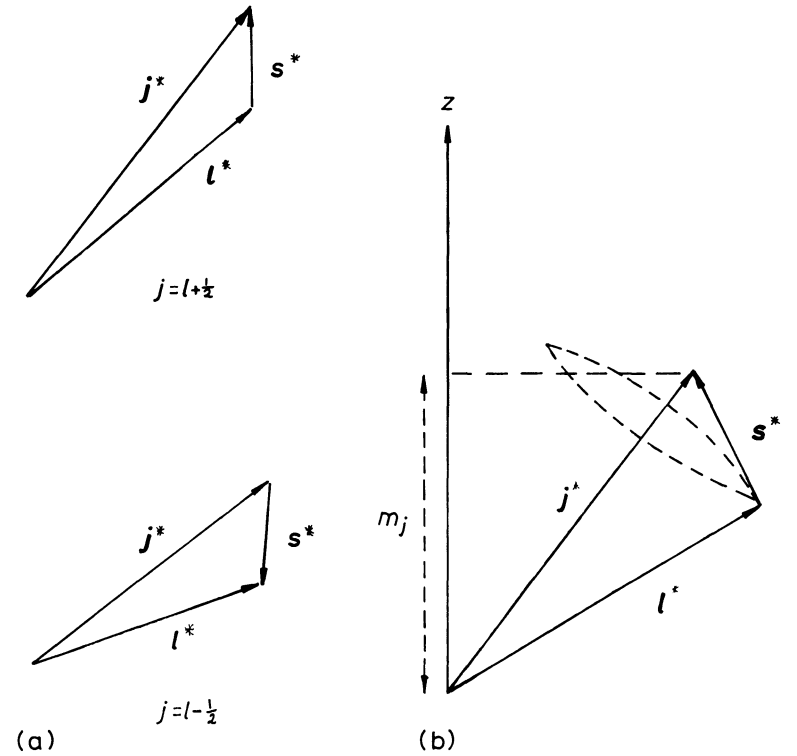

Figure 2.9 Magnetic spin–orbit interaction. As in Fig. 2.7, the lengths of the arrows represent the magnitudes of the vectors. (a) Vector model for the quantum number j. (b) Vector model for the quantum number m_j.

resultant j, so their projections on the z-axis are no longer constant (Fig. 2.9(b)). The state of the electron is now described by j and the projection of j on the z-axis, m_j. The quantum numbers n, l, j and m_j form an alternative set to n, l, m_l and m_s. In the absence of an external field a level of given j has a degeneracy of $2j+1$ corresponding to the possible values of m_j from $+j$ to $-j$. It can easily be verified that the total number of states of given l in this representation – i.e. $2(l+\frac{1}{2})+1$ from $j=l+\frac{1}{2}$ plus $2(l-\frac{1}{2})+1$ from $j=l-\frac{1}{2}$ – is identical with the number $2(2l+1)$ in the m_l, m_s representation.

In an atom with several valence electrons the total angular momentum $\boldsymbol{J}\hbar$ can be defined in a similar way by $\boldsymbol{J}=\boldsymbol{L}+\boldsymbol{S}$, provided the magnetic interaction is small compared with the electrostatic. The quantum number J can take any of the $(2S+1)$ integral or half-integral values $L+S, L+S-1, \ldots, L-S$. S thus determines the multiplicity of the term. If $L < S$ there are only $(2L+1)$ possible values of $J, S+L, \ldots, S-L$, but the superscript $2S+1$ is still used to describe the term. In particular, S-terms ($L=0$) are unsplit whatever their multiplicity.

The spin–orbit interaction energy has the same form as in the one-electron case:

$$\Delta E_{LS}=A\boldsymbol{L}\cdot\boldsymbol{S}=\tfrac{1}{2}A(J^{*2}-L^{*2}-S^{*2}) \tag{2.28}$$

The constant A can be determined in the simpler cases from the as of the individual electrons. As in the single electron case, the levels are labelled with J as a subscript. For example, a ^{4}D term ($L=2, S=\frac{3}{2}$) is split into four levels $^4D_{1/2}$, $^4D_{3/2}$, $^4D_{5/2}$ and $^4D_{7/2}$. These levels are not equally spaced: it is easily shown from equation 2.29 that the interval between two levels of a given term is proportional to the larger of the two J-values concerned, a relation known as the Landé interval rule.

Figure 2.10 shows the complete set of energy levels derived from the sp configuration considered in the previous section, with the threefold degeneracy

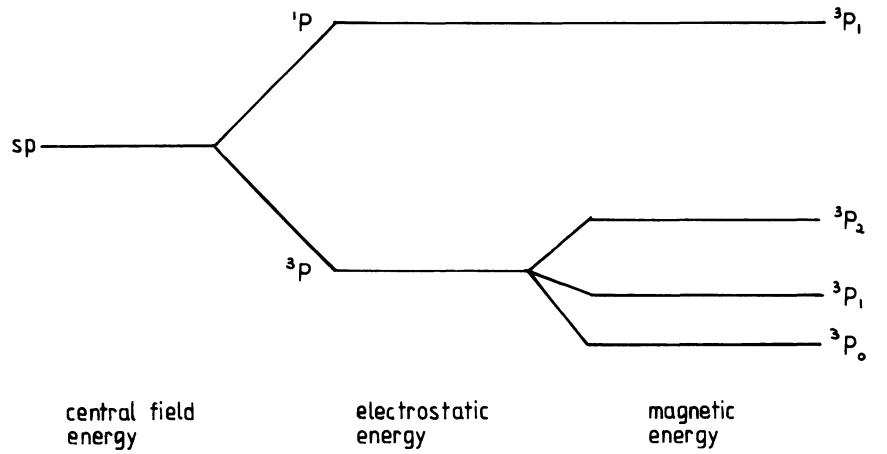

Figure 2.10 Electrostatic and magnetic (spin–orbit) interactions in the sp-configuration.

of ^{3}P removed by the spin–orbit interaction. Figure 2.11 shows the fine structure (not to scale) in (a) an atom with one valence electron (potassium) and (b) an atom with two valence electrons (calcium).

In the examples of Fig. 2.11 the fine structure is small, of order 0.01 eV, compared with the term separation, which is typically 1 eV. This is a necessary condition for the *L–S*, or Russell–Saunders coupling scheme just described, for otherwise L and S are not meaningful quantum numbers. The heavier the element is, the greater the fine structure; for example, the lowest P-term of cesium is split by 554 cm^{-1}, compared with 17 cm^{-1} for sodium. The *L–S* coupling scheme therefore breaks down for heavy elements, and at the other extreme a system may be described by *j–j* coupling, where $\boldsymbol{j}_1 = \boldsymbol{l}_1 + \boldsymbol{s}_1$, etc., and $\boldsymbol{J} = \boldsymbol{j}_1 + \boldsymbol{j}_2 + \ldots$. In general an intermediate coupling scheme may be required. In any event, J is always a 'good quantum number', and for a given electron configuration the total number of levels and the J-values are independent of the coupling scheme. It is the arrangement and grouping of these levels that are determined by the coupling. In addition to J, the other 'good quantum number' is M_J, the projection of J on the z-axis. This can take the $2J+1$ values from $+J$ to $-J$, and in the absence of an external field each J-value has a degeneracy of $2J+1$.

The hydrogen atom and the one-electron ions, He^+, Li^{2+}, etc., are special cases. The peculiarity of their l-degeneracy has already been mentioned, and in addition their fine structure is so small that other relativistic effects have to be

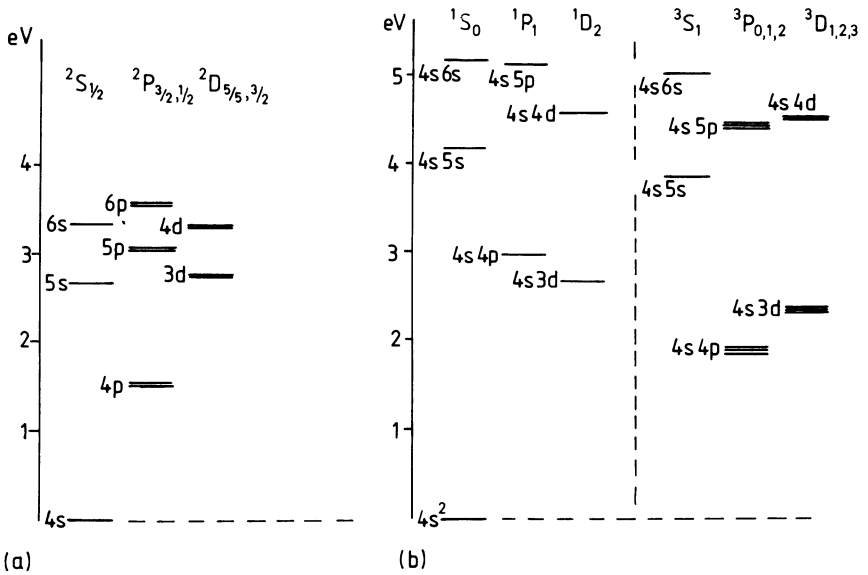

Figure 2.11 Energy level diagrams for (a) potassium and (b) calcium. The doublet and triplet splittings in these diagrams are not to scale.

treated at the same time. A full relativistic correction, incorporating the spin interaction, actually replaces l-degeneracy with j-degeneracy, as shown in Fig. 2.12. Although this looks superficially like an ordinary central field energy scheme, with n levels for quantum number n, the levels are characterized by j, not by l, and different selection rules apply to transitions between them. The fine structure correction is given by

$$\Delta E_{n,j} = E_n \alpha^2 \frac{Z^2}{n^2}\left(\frac{3}{4} - \frac{n}{j+\frac{1}{2}}\right) \tag{2.29}$$

where $Z=1$ for H, 2 for He^+, etc., and α is the fine structure constant,

$$\alpha = \frac{e^2}{(4\pi\varepsilon_0)\hbar c} \simeq \frac{1}{137}$$

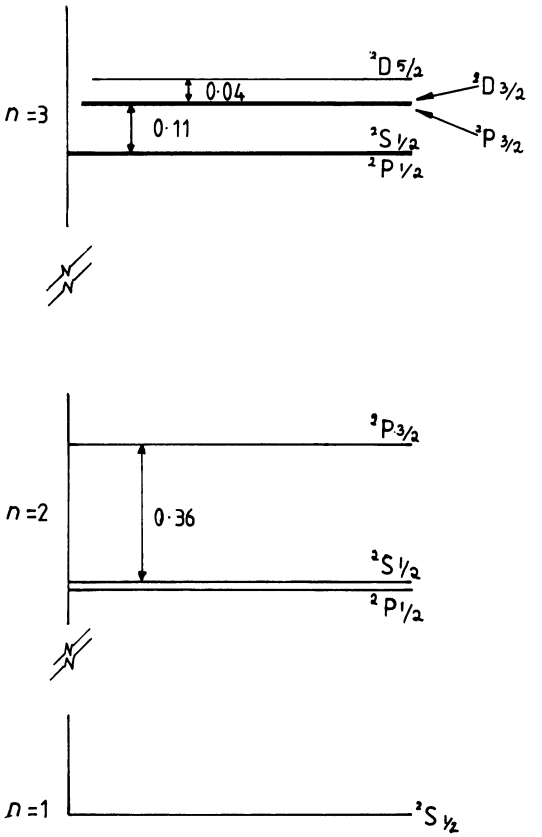

Figure 2.12 Hydrogen fine structure. The splittings shown are in cm^{-1}. The $^2S_{1/2}$ and $^2P_{1/2}$ levels are not quite degenerate because of the Lamb shift, which is 0.03 cm^{-1} for $n=2$ and decreases rapidly with increasing n.

α is one of the fundamental dimensionless constants of physics. Its relativistic origin can be seen from the relation $\alpha = v/c$, where v is the speed of the electron in the first Bohr orbit. Equation 2.29 is correct to terms in v^2/c^2. The splitting in the $n=2$ state is only 0.36 cm^{-1} in H, and as is clear from equation 2.29 it decreases rapidly with increasing n.

Actually, this relativistic correction is not the end of the hydrogen story. Very small discrepancies between theory and experiment led to the discovery of a small shift of $s_{1/2}$ relative to $p_{1/2}$, known as the Lamb shift, an order of magnitude smaller than the fine structure splitting. A quantum-dynamical treatment is needed to explain this shift. It can be interpreted as the interaction of the electron with the fluctuating electric field due to the zero-point energy of the quantized radiation field and can be expressed theoretically in powers of $Z\alpha$. Comparison of the theoretical and experimental values of this very small shift, which is appreciable only for s-states, offers a sensitive test of quantum electrodynamical theory; some of the experimental methods used are described in Chapters 8 and 9.

2.8 Radiative transitions and selection rules

Radiative transitions are concerned with the interaction between an atom and a radiation field, and a full treatment requires quantum electrodynamics. The essential results for our purpose can be obtained by treating the electromagnetic field as a time-dependent perturbation. Section 2.5 gives the result of such a perturbation as a probability $|c_k(t)|^2$ of finding the electron in a final state u_k different from its initial state u_i. The perturbation $\mathbf{H}^1(t)$ to be inserted in equation 2.23 is the interaction of the atom with the electric and magnetic fields $\boldsymbol{E} = \boldsymbol{E}_0 \cos(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})$ and $\boldsymbol{B} = \boldsymbol{B}_0 \cos(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})$ of the electromagnetic wave. This interaction can be expanded as a series in order of decreasing strength: electric dipole, magnetic dipole, electric quadrupole, etc. Initially we keep just the first term, which amounts to assuming the electric field to be constant over the atomic dimensions ($\lambda \gg a_0$), an assumption justified down to the X-ray region since $a_0 \sim 1$ Å. As we shall see, the higher-order terms may be important if the electric dipole interaction vanishes.

Suppose we have a plane wave polarized in the x-direction, for which the electric field contributed by radiation between ω and $\omega + d\omega$ is $E_\omega \cos \omega t$. The ω-component of the electric dipole interaction is then $ex \cdot E_\omega \cos \omega t$. Writing $(E_k - E_i)/\hbar = \omega_{ik}$, equation 2.23 becomes

$$[c_k(t)]_\omega = -\frac{i}{\hbar} E_\omega \langle k|ex|i\rangle \int_0^t dt\, e^{i\omega_{ik}t} \cos \omega t \tag{2.30}$$

The next step is to put $\cos \omega t = (e^{i\omega t} + e^{-i\omega t})/2$ and perform the t-integration. This produces a term $(\omega_{ik} \pm \omega)$ in the denominator which ensures that all c_k are

vanishingly small except the particular one, c_f, for which ω_{if} is close to ω:

$$[c_f(t)]_\omega = -\frac{1}{\hbar}\frac{E_\omega}{2}\langle k|ex|i\rangle\left(\frac{e^{i(\omega_{if}-\omega)t}-1}{\omega_{if}-\omega}+\frac{e^{i(\omega_{if}+\omega)t}-1}{\omega_{if}+\omega}\right)$$

For $\omega_{if}\simeq\omega$ we can ignore the second term on the right. Multiplying by the complex conjugate, we have

$$|c_f(t)|^2_\omega = \frac{E_\omega^2}{4\hbar^2}|\langle k|ex|i\rangle|^2\frac{4\sin^2\frac{1}{2}(\omega_{if}-\omega)t}{(\omega_{if}-\omega)^2} \tag{2.31}$$

The final step is to integrate over ω. In fact, $c_f(t)$ is effectively zero except for the narrow range of ω around ω_{if}, so E_ω in equation 2.31 can be regarded as a constant, $E(\omega_{if})$. The reason for taking the ω-integration over the probability $|c_f|^2$ rather than the probability amplitude c_f is that contributions from different frequencies are assumed to add incoherently. Since $E(\omega_{if})$ is related to the radiation density $\rho(\omega_{if})$ by $\rho=\frac{1}{2}\varepsilon_0|E|^2$ for plane-polarized light, we have after evaluation of the (standard) integral:

$$|c_f(t)|^2 = \frac{E^2(\omega_{if})}{2\hbar^2}|\langle k|ex|i\rangle|^2\pi t = \frac{\pi\rho(\omega_{if})}{\varepsilon_0\hbar^2}|\langle k|ex|i\rangle|^2 t$$

For isotropic unpolarized light, the y- and z-components make equal contributions to the dipole interaction, and the energy density is divided equally among the three components. The final result for the transition probability *per unit time* is:

$$P_{if} = \frac{|c_f(t)|^2}{t} = \frac{\pi\rho(\omega_{if})}{3\varepsilon_0\hbar^2}S_{if} \tag{2.32}$$

where $\rho(\omega_{if})$ is the radiation density *per unit interval of angular frequency* and the electric dipole line strength S_{if} is defined by

$$S_{if} = |\langle f|er|i\rangle|^2 = |\langle f|ex|i\rangle|^2 + |\langle f|ey|i\rangle|^2 + |\langle f|ez|i\rangle|^2 \tag{2.33}$$

The effects of degeneracy and the relations between absorption and emission transition probabilities will appear in sections 4.5 and 11.3. In this section we shall introduce the selection rules that determine whether P_{if} is non-zero – i.e. whether the transition is 'allowed'.

These selection rules can be discussed with reference to $\langle f|ex|i\rangle$, as similar considerations apply to the other two components. By writing $x=r\sin\theta\cos\varphi$ and $u=R(r)\,Y(\theta,\varphi)$ (equation 2.12), the integral can be expressed as the product of two integrals, one involving r only and the other angles only. The Y-function for any electron in a central field is identical with that of hydrogen, as we have already seen. Under the inversion $x\rightarrow -x$, which requires $r\rightarrow r$, $\theta\rightarrow\theta$ and $\varphi\rightarrow\pi-\varphi$, Y has the property that $Y\rightarrow(-1)^l Y$. Since R is unchanged by the inversion, we have $u\rightarrow +u$ for even l and $u\rightarrow -u$ for odd l. If the positive and negative contributions to $\int xu_f^* u_i\,d\mathbf{r}$ are not to cancel, the product $u_i u_f^*$ must

change sign on inversion, and this evidently requires either Y_i or Y_f to have odd l; in other words, their l-values must differ by 1, 3, 5, ... It can actually be shown that the Y-integral vanishes except when the difference is 1, and the first selection rule becomes

$$\Delta l = \pm 1 \tag{2.34}$$

In particular, electric dipole transitions between terms of the same configuration are forbidden.

This selection rule can be generalized by introducing the concept of parity. A term is said to have odd or even parity according to whether its wave function does or does not change sign on inversion. The function for several independent electrons is simply the product of the individual functions, and the parity is therefore determined by $(-1)^{\Sigma l}$ and is odd or even as the algebraic sum Σl is odd or even. Obviously all terms of a single configuration have the same parity. The significance of parity is that it remains well defined even when interaction effects make the ls of the individual electrons no longer meaningful. The Δl selection rule is then replaced by Laporte's rule: electric dipole transitions occur only between terms of opposite parity.

We have seen that the J-value of any term is also well defined, and there is a corresponding selection rule

$$\begin{gathered}\Delta J = 0, \pm 1, \quad J = 0 \nleftrightarrow J = 0 \\ \text{(i.e. } J = 0 \rightarrow J = 0 \text{ forbidden)}\end{gathered} \tag{2.35}$$

This rule cannot be derived from the Schrödinger wave functions, which do not include spin. Physically it comes from conservation of angular momentum, the spin angular momentum of the photon being $\hbar$; one of the references [1–6] should be consulted for the derivation of this selection rule.

Insofar as L and S are good quantum numbers, there are additional selection rules

$$\begin{gathered}\Delta L = 0, \pm 1, \quad \text{with } L = 0 \rightarrow L = 0 \text{ forbidden} \\ \text{and} \qquad \Delta S = 0\end{gathered} \tag{2.36}$$

These rules depend on the validity of the L–S coupling scheme, and do not hold well for the heavier elements, in which the 'intercombination' lines between terms of different S can be quite strong.

If a given transition is forbidden by the electric dipole selection rules, it may still occur by magnetic dipole or electric quadrupole transitions, which come from the next terms in the expansion of the interaction of the electromagnetic field with the atom. The corresponding transition probabilities involve even powers of x, y and z and therefore obey the parity selection rule even→even or odd→odd. In particular, transitions between terms in the same configuration are allowed. The transition probabilities are far smaller than for electric dipole: the factor depends on wavelength, and in the visible region it is of order 10^{-5} for

magnetic dipole and 10^{-8} for electric quadrupole (see Chapter 11). The upper levels of such 'forbidden' lines are described as metastable. The lines are important in astrophysics because at sufficiently low gas densities a forbidden transition has a good chance of occurring before the atom is knocked out of its excited state by a collision. The frequency dependence of electric quadrupole transitions makes them increasingly important at very short wavelengths (when the approximation $\lambda \gg a_0$ begins to break down). Conversely, magnetic dipole transitions are important at radiofrequencies, and much of Chapter 9 is concerned with them.

It is the selection rules that form the link between observed spectra and energy level diagrams such as that of Fig. 2.11. Electric dipole transitions connect levels in adjacent columns in these diagrams: there are also weaker links between the singlet and triplet sets of levels in the two-electron system. Transitions from the ground state of potassium (Fig. 2.11(a)) are limited to the resonance line 4s–4p, the higher members of the so-called principal series 4s–*n*p and the continuum. The other important series are the sharp, 4*p*–*n*s, and the diffuse, 4*p*–*n*d. In the two-electron spectrum (Fig. 2.11(b)) each of these series appears twice, as a singlet series and a triplet series with different limits, and there are also intercombination lines such as $4s^2\ {}^1S$–$4s4p\ {}^3P$. In a complex spectrum the number of possible series is much greater; the transition pp–pd, for example, is made up of five different singlet series and five different triplet series, along with any intercombination lines.

In general each 'line' is split into a number of components by the fine structure. The potassium resonance line, for example, has two components

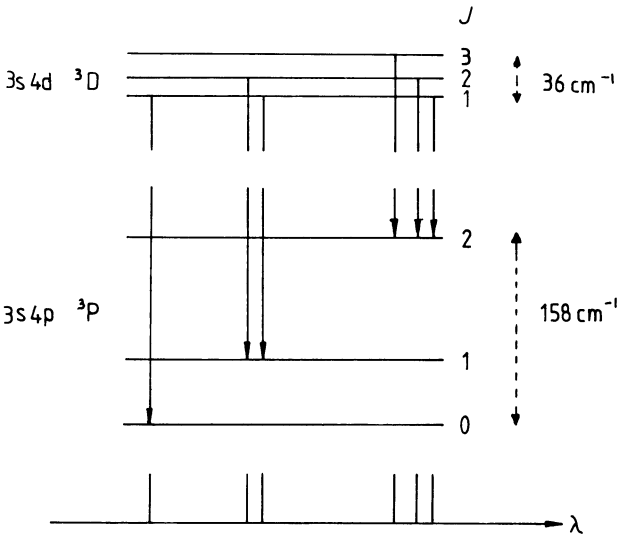

Figure 2.13 Complex triplet in the spectrum of calcium. Six components are allowed by the selection rules $\Delta J = 0, \pm 1$.

$4s\,^2S_{1/2}$–$4p\,^2P_{1/2}$ and $4s\,^2S_{1/2}$–$4p\,^2P_{3/2}$, separated by the width of the $4p\,^2P$ fine structure, 58 cm^{-1} or 3.4 nm. The group of lines arising from transitions between two particular terms is known as a multiplet. Figure 2.13 shows an example of a complex multiplet (both terms split) in the calcium spectrum of Fig. 2.11(b). In this case the total spread of the line pattern is about 160 cm^{-1}, and the splitting of ^3D is much smaller than that of ^3P, so the lines occur in three groups. Fine structure splittings decrease rapidly with increasing n and l in any given atom. Except in very light atoms the resolution of the fine structure in the lower series members does not usually present a problem. In heavy atoms the problem is rather to link together components of the same multiplet that may be separated by many tens of nm. In mercury the ^3P–^3D multiplet analogous to the calcium multiplet in Fig. 2.13 extends over 6200 cm^{-1}.

In complex spectra, particularly those involving partly filled d- and f-shells, such as the transition elements and the rare earths, the number of terms becomes vastly greater, and the allowed transitions give rise to an extremely high density of lines in certain spectral regions. For example, in parts of the iron spectrum there can be 100 or more lines in a 10 nm wavelength interval.

2.9 Zeeman and Stark effects

The Zeeman and Stark effects describe the splitting of energy levels, and hence spectral lines, in magnetic and electric fields, respectively. The Zeeman effect is much the easier to interpret and has been widely used to help to analyse spectra in the optical region. It plays a leading part in microwave and radiofrequency spectroscopy, as Chapter 9 shows. Historically it was the anomalous Zeeman effect that led Goudsmit and Uhlenbeck to postulate the existence of electron spin. The Stark effect is of little use in the interpretation of spectra, but it plays an important part in molecular spectroscopy and is responsible for pressure broadening by charged particles (Chapter 10).

The *Zeeman effect* can be described by a semi-classical model of the interaction of the resultant electronic magnetic moment with the external magnetic field. The intrinsic magnetic moment associated with electron spin has already been introduced to account for fine structure. If L and S are good quantum numbers, equation 2.26 can be written for several electrons as

$$\boldsymbol{\mu}_s = -g_s \mu_B \boldsymbol{S}^* \tag{2.37}$$

where $g_s = 2$, μ_B is the Bohr magneton and $\boldsymbol{S}^*$ has the magnitude $\sqrt{[S(S+1)]}$. Similarly, the magnetic moment associated with the orbital part of the angular momentum is

$$\boldsymbol{\mu}_L = -g_L \mu_B \boldsymbol{L}^* \tag{2.38}$$

Both experimentally and theoretically g_L is unity: that is, the gyromagnetic ratio for orbital angular momentum has the classical value $e/2m$. However, $\boldsymbol{L}^*$ and $\boldsymbol{S}^*$

both precess about their resultant $\boldsymbol{J}^*$, as shown in Fig. 2.9 for the one-electron case, and so the components of $\boldsymbol{\mu}_S$ and $\boldsymbol{\mu}_L$ perpendicular to $\boldsymbol{J}^*$ average to zero. Only the components in the $\boldsymbol{J}^*$-direction contribute to the resultant magnetic moment μ_J:

$$\mu_J = \mu_L \cos(L, J) + \mu_S \cos(S, J)$$

$\boldsymbol{\mu}_J$ is thus parallel to $\boldsymbol{J}^*$ and can be expressed as

$$\boldsymbol{\mu}_J = -g_J \mu_B \boldsymbol{J}^* \tag{2.39}$$

This equation defines the Landé g-factor g_J, which can be evaluated for L–S coupling by applying the cosine rule to the angles (L, J) and (S, J):

$$g_J = 1 + \frac{J^{*2} + S^{*2} - L^{*2}}{2J^{*2}} \tag{2.40}$$

Evidently $g_J = 1$ for a singlet term ($S = 0$, $J = L$). For an S-term ($L = 0$), in which there is no orbital contribution to the magnetic moment, $g_J = 2$. If L–S coupling does not hold, g_J is still defined by equation 2.39 but cannot be evaluated from equation 2.40; the g_J value is, in fact, a good diagnostic of the type of coupling.

The energy of the magnetic moment μ_J in the field $\boldsymbol{B}$ is

$$\Delta E_B = -\boldsymbol{\mu}_J \cdot \boldsymbol{B} = g_J \mu_B \boldsymbol{J}^* \cdot \boldsymbol{B}$$

B defines the z-axis, and $\boldsymbol{J}^* \cdot \boldsymbol{B}$ is therefore equal to $M_J B$ where M_J is restricted to the values $J, J-1, \ldots, -J$. Therefore,

$$\Delta E_B = g_J \mu_B B M_J \tag{2.41}$$

The magnetic field splits each level into $2J + 1$ sub-levels of equal separation $g_J \mu_B B$, as shown in Fig. 2.14 for $J = \frac{3}{2}$. This removes the last degeneracy in the wave function: each magnetic sub-level corresponds to one and only one set of quantum numbers.

The selection rule for electric dipole radiation is

$$\Delta M_J = \begin{cases} \pm 1, & \sigma \text{ components} \\ 0, & \pi \text{ components} \end{cases} \tag{2.42}$$

Viewed perpendicularly to the field, the σ- and π-components have the electric vector of the electromagnetic wave polarized parallel and perpendicular, respectively, to B; viewed along the B-direction, the σ-components are circularly polarized and the π-components do not appear at all. Figure 2.15 shows the Zeeman pattern for (a) a $^2S_{1/2}$–$^2P_{3/2}$ transition – e.g. one of the sodium yellow lines – and (b) a $J = 1$–$J = 2$ transition – e.g. one component of the ^3P–^3D multiplet of Fig. 2.13. If the g_J factors of the two levels are exactly equal – in particular, if they are both unity, as in a singlet–singlet transition – the spacings in the upper and lower levels are also equal. All the $\Delta M_J = +1$ transitions then coincide, as do the $\Delta M_J = 0$ and $\Delta M_J = -1$ transitions, giving only three

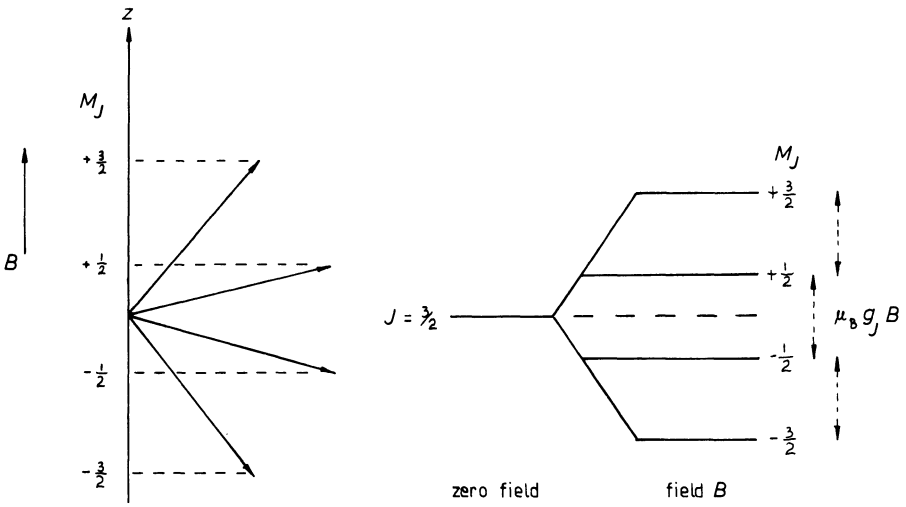

Figure 2.14 Splitting of a $J=\frac{3}{2}$ level in an external magnetic field B.

Zeeman components separated by $\mu_B B$. This is the so-called 'normal' Zeeman effect; it is strictly abnormal in that it occurs only for singlet states. If the g-factors are not very different, as in Fig. 2.15, then the components still form σ–π–σ groups; but if the gs are very different the pattern bears little resemblance to the 'normal' triplet and may be very complicated for large J. It is always symmetric about the field-free line.

As g_J is not very different from unity, the order of magnitude of the splitting is determined by $\mu_B B$, which is $0.046\ \text{cm}^{-1}$ for a field of 0.1 T (1 kG). This is comparable with the Doppler linewidth for a light element (section 10.3), and it is necessary to use high fields, high resolution and light sources with narrow lines to resolve a complex Zeeman pattern. Magnetic dipole transitions between the sub-levels of the same J-level (Chapter 9) provide the most accurate determinations of the splittings.

In this treatment it has been implicitly assumed that the interaction with $\boldsymbol{B}$ is small compared with the internal interactions in the atom, so spin–orbit coupling is undisturbed and J remains a good quantum number. If the field is high enough for the Zeeman splitting to compare with the fine structure, which may be only about $1\ \text{cm}^{-1}$ for a light element, we have to think about the validity of this assumption. To go to the opposite extreme, if it were possible to make $\mu_B B$ much greater than the L–S interaction, L and S would interact directly with the field, with $(M_L + 2M_S)\,\mu_B B$ as the principal term in the energy and the L–S interaction energy as a small correction. This limit is known as the Paschen–Back effect. It is not of much practical importance as far as fine structure is concerned because the fields required are so large, but the analogous

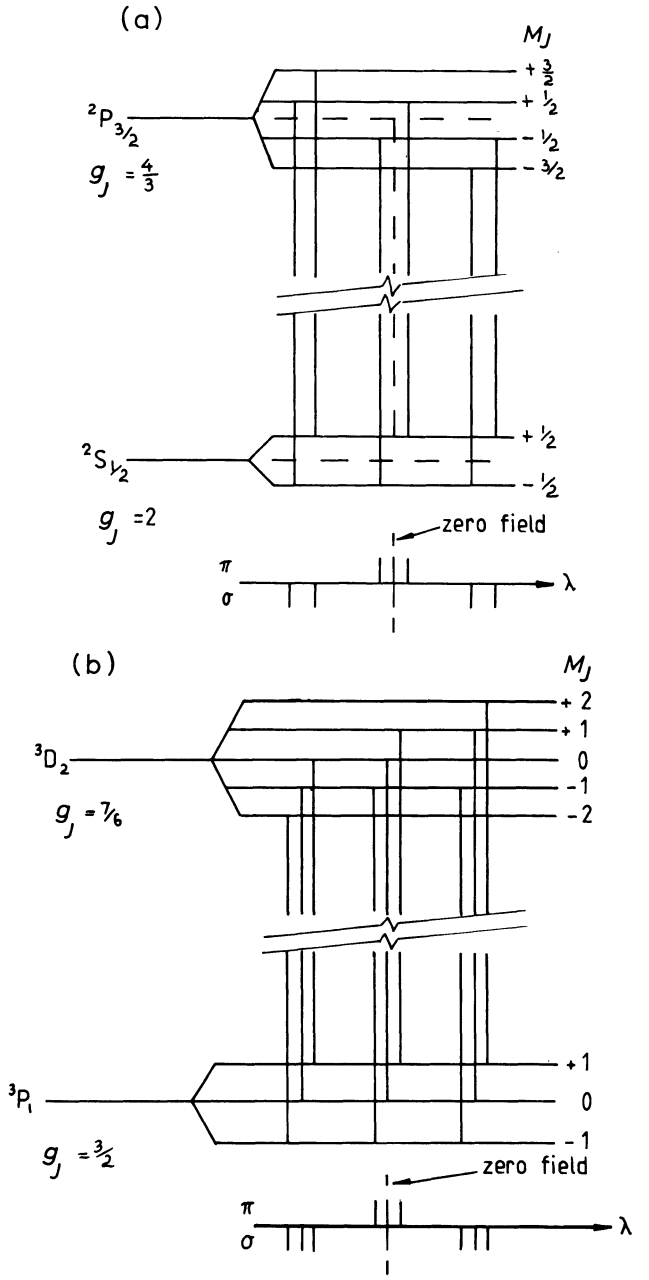

Figure 2.15 Zeeman patterns for the transitions (a) $^2S_{1/2}$–$^2P_{3/2}$ and (b) $^3P_1$–$^3D_2$. In both cases the g-factors of the upper and lower states are not very different, and the components form three groups corresponding to the three components of a 'normal' Zeeman pattern. σ-Components obey $\Delta M_J = \pm 1$ and π-components $\Delta M_J = 0$.

effect in hyperfine structure *is* important and is studied mainly by the methods of Chapter 9. The intermediate-field case, in which the internal and external magnetic interactions are of comparable size, is important for both fine and hyperfine structure, and we return to this in Chapter 9.

The *Stark effect* differs from the Zeeman effect in that atoms have no permanent electric dipole moment to interact with an electric field. It cannot be treated in a simple semi-classical fashion. We shall outline here the perturbation treatment, with the aim of establishing a base for the description of pressure broadening in Chapter 10.

In all atoms except hydrogen, which must be treated separately because of its l-degeneracy, the effect of lowest order is quadratic: the interaction energy is proportional to the square of the electric field, $|E|^2$. One can think of the electric field as distorting the electron distribution so as to induce a dipole $\alpha \boldsymbol{E}$, where α is the polarizability. The interaction between this induced dipole and the field is αE^2. α may reasonably be expected to depend on the angular distribution of the electron charge, and hence on M_J. The distributions represented by $+M_J$ and $-M_J$ have the same symmetry with respect to the z-axis, which is defined by the direction of $\boldsymbol{E}$, so the Stark interaction depends only on $|M_J|$, and each sub-level except $M_J = 0$ remains doubly degenerate.

The perturbation operator is given by

$$\mathbf{H}^1 = e\boldsymbol{r} \cdot \boldsymbol{E} = ezE$$

but the first-order perturbation energy evaluated from the unperturbed wave functions u_i is zero because $\langle i|ez|i\rangle = 0$, as seen in section 2.8 in connection with the selection rule $\Delta l = 1$. This is just another way of saying that the atom has no permanent dipole moment. We therefore go on to the second-order perturbation, using the perturbed wave functions given in equations 2.19 and 2.20 to evaluate the interaction energy:

$$u_i^1 = u_i + \sum_{k \neq i} c_k u_k \quad \text{with} \quad c_k = \frac{\langle k|ez|i\rangle}{E_i - E_k} E \tag{2.43}$$

The integral is identical with the z-component of the electric dipole transition probability for states i and k (equation 2.33). Hence, the rules that determine which states k are to be mixed with state i are identical with the selection rules for electric dipole transitions. The contribution to the interaction energy made by each state for which c_k does not vanish is given by equation 2.21:

$$\Delta E_i = c_k \langle i|ezE|k\rangle = \frac{|\langle k|ez|i\rangle|^2}{E_i - E_k} E^2 \tag{2.44}$$

From the sign of the denominator it can be seen that the state i is effectively pushed away from the perturbing state k, and from the symmetry of the expression the state k is pushed away an equal amount by i. The electric field thus causes the two states to repel one another by an amount proportional to E^2

and inversely proportional to their separation. Evaluation of the integrals shows that the repulsion also depends on M_J^2. The 'selection rules' for the perturbation are the 'opposite parity' rule together with $\Delta J = 0, \pm 1$ and $\Delta M_J = 0$ (the $\Delta M_J = \pm 1$ contributions vanish since the electric field is by definition along the z-axis).

Figure 2.16 illustrates the physical basis of the perturbation. The external field derives from the potential $V_E(z)$, which is superposed on the symmetric central-field potential $V_0(r)$ to give a resultant V' that is asymmetric in the z-direction. The symmetric unperturbed wave function is given the necessary asymmetry to fit V' by mixing in a function of the opposite parity. The mixing of states of different l means that l is no longer a 'good quantum number' for the perturbed state; this is to be expected since the θ-dependence of V' prevents the separation of the Schrödinger equation into independent radial and angular parts. M remains a good quantum number because the potential is still cylindrically symmetric.

Figure 2.17 shows the Stark splitting of the alkali resonance lines as an example of the quadratic Stark effect. The ground state is not much perturbed because the first excited configuration is rather a long way above it. Since the density of levels increases as the excitation increases, levels of moderate excitation tend to be shifted downwards, giving a 'red shift' to the transition as a whole. The $|M_J|$ dependence results in an asymmetric splitting of the shifted line.

Fields in the range 10^6–10^7 V m^{-1} are usually required to produce observable Stark splitting – say a few tenths of a wavenumber. The splitting increases with

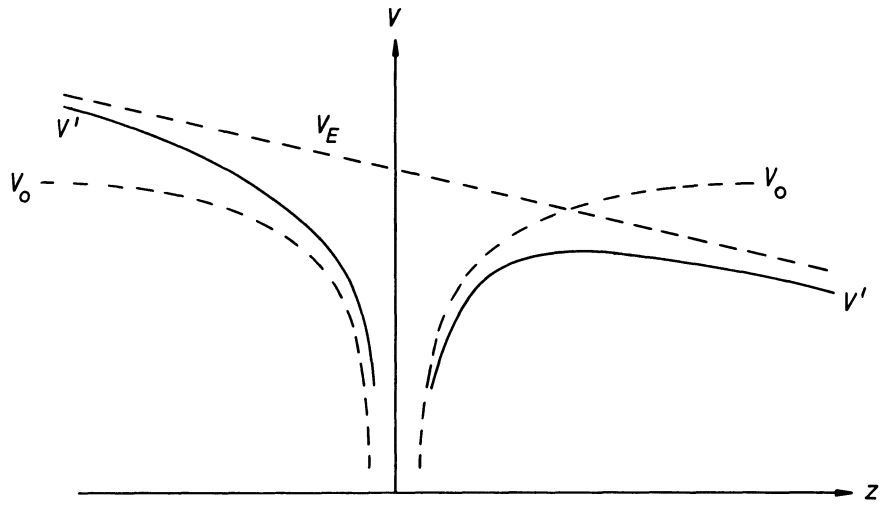

Figure 2.16 Distortion of the central field potential V_0 by the external electric field potential V_E (broken curves). The full curves V' show the resultant asymmetric potential.

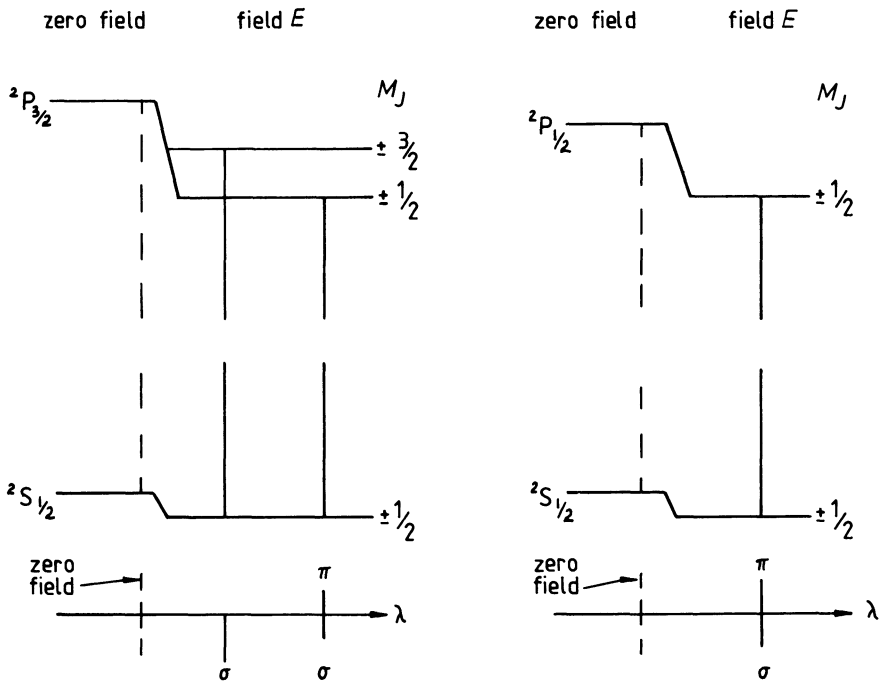

Figure 2.17 Quadratic Stark effect for the alkali resonance lines. Both lines are shifted to the red, and the $^2S_{1/2}$–$^2P_{3/2}$ line is also split.

the excitation because of the greater density of states. At values of the principal quantum number n so high that Stark splitting becomes comparable with the energy differences between states of different l, the latter become effectively degenerate, the second-order perturbation treatment breaks down, and the Stark effect becomes linear in E, as with hydrogen.

Another consequence of the l mixing induced by the field is the breaking of the dipole radiation selection rule $\Delta l = \pm 1$. For example, the admixture of p- in the s-wave function allows the 'forbidden' transitions s–d and s–s to appear in the alkali spectra in the presence of an electric field.

The hydrogen atom shows a linear Stark effect because of its l-degeneracy. The fine structure from spin–orbit and relativistic effects is of order 0.1 cm^{-1}, which is comparable with the Doppler linewidth at room temperature, so it may usually be ignored compared with a Stark splitting large enough to be readily measurable. All states of the same n are thus effectively degenerate. In attempting to apply perturbation theory to degenerate states one immediately runs into the difficulty that the c_k in equation 2.20 become infinite unless the numerator is also zero – i.e. we must have

$$\langle k|ez|i\rangle = 0 \quad \text{when} \quad E_k = E_i$$

This condition is not satisfied when k and i are two states of the same energy but different l. The integral can be made to vanish by using for the zero-order functions some linear combination of u_i and u_k, say u_i' and u_k', where $u_i' = a_1 u_i + b_1 u_k$ and $u_k' = a_2 u_i + b_2 u_k$. The new functions satisfy the unperturbed Schrödinger equation with the same energy E^0 since linear combinations of solutions are themselves solutions. However, the *first*-order perturbation energy (equation 2.18) evaluated with these new functions no longer vanishes, because by mixing functions of opposite symmetry we have necessarily made them asymmetric:

$$\int u_i'^* \, ez E u_i' \, \mathrm{d}\boldsymbol{r} \neq 0 \quad \text{and} \quad \int u_k' \, ez E \, u_k' \, \mathrm{d}\boldsymbol{r} \neq 0$$

There is therefore first-order Stark splitting proportional to E.

The important distinction from the quadratic effect is that the linear combination of degenerate functions does not contain coefficients proportional to E. One set of zero-order functions is just as permissible as another until some perturbation forces one to choose a physically sensible combination. A similar situation has already been discussed in the case of two valence electrons, when the introduction of the electrostatic repulsion necessitated a choice of wave functions compatible with the indistinguishability of the electrons.

The result of the full treatment for hydrogen is that a state of given n splits symmetrically into $2n-1$ sub-levels, whose separation is proportional to E and to n. The line pattern is therefore also symmetric about the field-free line. Figure 2.18 shows the effect for the Balmer α-line, $n=2 \leftrightarrow n=3$. The splitting is a few cm^{-1} in a field of order $10^5 \ \mathrm{V\,m}^{-1}$, which is very much larger than the quadratic effect shown by other atoms. The Stark splitting of hydrogen is of very great importance in plasma spectroscopy because hydrogen is a common constituent of most plasmas and its Stark pattern can be calculated with reliability. Theoretical and experimental line shapes from the fields due to the charged particles in the plasma can be accurately compared, and as a result electron densities can be deduced from measured hydrogen line profiles.

2.10 Configuration interaction and auto-ionization

So far it has been assumed that an atom in zero external field has a single well-defined configuration so that a definite value of n and l can be allocated to each electron on the basis of the central field model. A comparison of theoretical with experimental values of energy levels and transition probabilities shows that this assumption is not always justified. The system may be better described by a different set of basis functions, formed by mixing into the original functions contributions from one or more other configurations. Without an external field to make the potential asymmetric, we can mix only functions of the same parity, in contrast with the Stark effect. The mixing coefficients c_k in equation 2.20 are zero unless the states i and k have the same parity and the same J-value. c_k falls

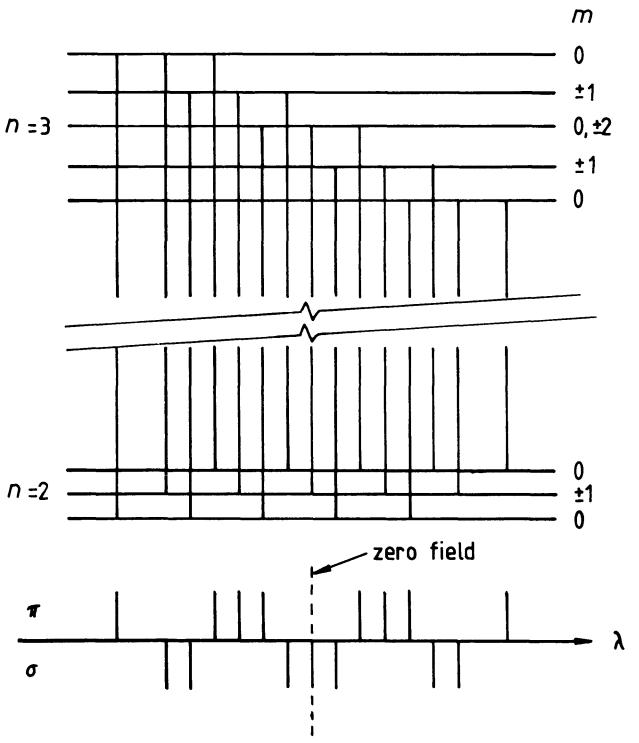

Figure 2.18 Linear Stark effect for the Balmer H_α line, showing symmetric splitting.

off inversely as the energy difference $|E_i - E_k|$, so that for most simple spectra the mixing is relatively unimportant; but in complex spectra, where there are several valence electrons, several states may easily have comparable energies. For example, in the transition elements the excitation energies of the 3d- and 4s-electrons are very similar and the parity condition for mixing is satisfied. Even in 'simple' spectra configuration interaction can be significant if two valence electrons are excited. In Be, for which the ground state is $2s^2\ {}^1S_0$, the 'anomalous' term $2p^2\ {}^3P$ has about the same energy as the terms arising from the 2s3s and 2s3d configurations of the normal spectrum. Mixing is therefore to be expected between levels of the same J. $2p^2$ does not mix with 2s3p, because of the parity rule.

Of particular interest is the configuration mixing between a bound state and an adjacent ionization continuum that gives rise to auto-ionization. In a simple term diagram (e.g. Fig. 2.11) only one valence electron is excited and all terms converge to the same limit, corresponding to the ion in its ground state plus the electron with zero kinetic energy. In the states of the continuum beyond the series limit the electron has finite kinetic energy. However, there exist bound states equal in energy to the continuum states, due either to the simultaneous

excitation of two electrons or to the excitation of an electron from an inner shell – e.g. one of the p-electrons from the p^6 closed shell in an alkali atom. Figure 2.19 shows, as an example of the first case, the energy level diagram for calcium, with the singlet–triplet splitting omitted for simplicity. On the left-hand side are the 'normal' terms 4s*nl*, converging to the 4s ground state of Ca^+. On the right-hand side are the 'anomalous' terms, in which one of the valence electrons is in either a 3d or a 4p state while the other is excited to various higher states. Removal of this second electron from the atom leaves the calcium ion in either a 3d ^{2}D or a 4p ^{2}P state. Almost all of the bound states converging to either of these two limits have greater energy than the normal ionization energy and can mix with the continuum states of the same parity and energy. The bound state becomes to some extent part of the continuum and is thereby broadened. One can also think in terms of the mixing representing a certain probability of finding the system in the ionized state – in other words, of

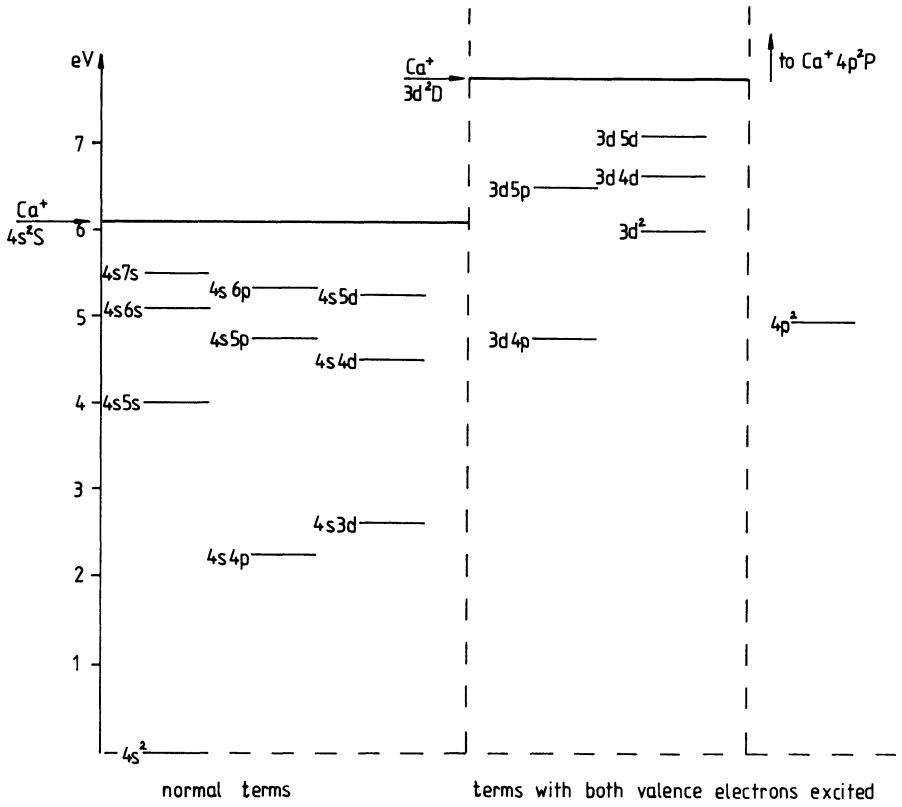

Figure 2.19 Energy level diagram for calcium, showing on the left the normal terms converging to the ground state of Ca^+ and on the right the 'anomalous' terms converging to excited states of Ca^+. The term splitting has been omitted for simplicity.

undergoing a radiationless transition as shown schematically in Fig. 2.20. This process is known as *auto-ionization.* Its probability may be as high as 10^{13} s^{-1}, compared with a typical value of 10^8 s^{-1} for an allowed radiative transition. It will be seen in section 10.2 that the natural width of a line depends on the lifetimes of the levels it connects according to the uncertainty principle $\Delta\nu\Delta t \sim 1$. For an auto-ionizing level Δt may be several orders of magnitude smaller than for a normal bound level, and an absorption line with an auto-ionizing upper level is very diffuse, possibly as much as a few thousand wavenumbers wide. Anomalous levels that do not have the right parity to auto-ionize have normal widths.

Absorption lines from the ground state ending on auto-ionizing levels are necessarily in the far ultra-violet because they lie beyond the normal series limits. Study of the lines has shown them to be very asymmetric as well as wide. The profiles were first calculated by Fano, and Fig. 2.21 shows an example. The absorption at some distance from the line is the ordinary continuous absorption above the normal series limit. The auto-ionizing transition forms a resonance in this continuum, typically rising steeply on one side to a peak value and dropping on the other almost to zero before going back to the steady value. The absorption on this side of the line is thus *less* than that of the unperturbed

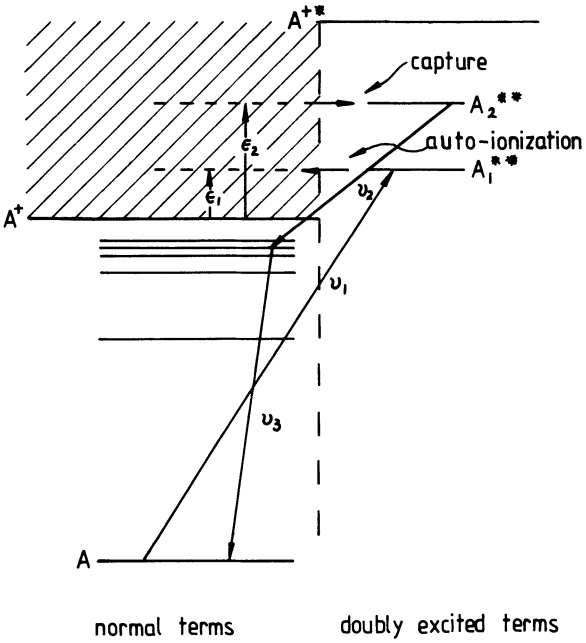

Figure 2.20 Schematic representation of auto-ionization and dielectronic recombination.
Auto-ionization: $A + h\nu_1 \rightarrow A_1^{**} \rightarrow A^+ + e^-(\varepsilon_1)$
Dielectronic recombination: $A^+ + e^-(\varepsilon_2) \rightarrow A_2^{**} \rightarrow A^* + h\nu_2 \rightarrow A + h\nu_3$

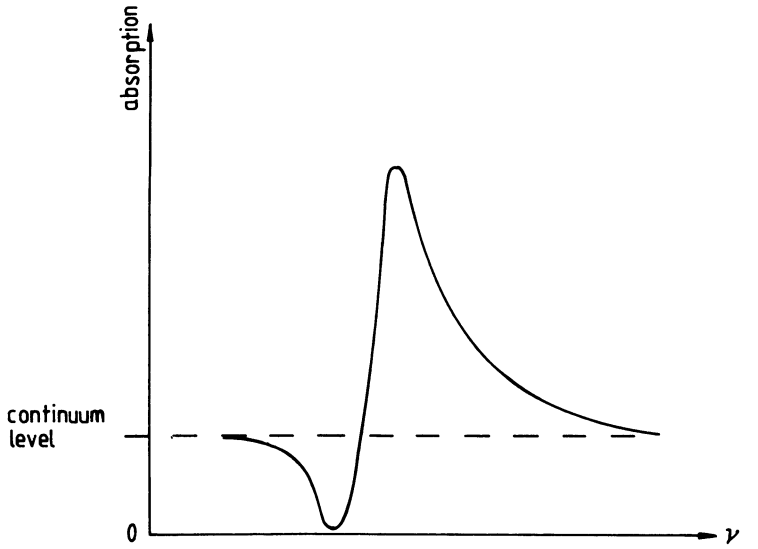

Figure 2.21 Example of asymmetric auto-ionizing line profile. The dip on the left appears as a window in the continuous absorption.

continuum and appears as a window in the latter. The 'explanation' of this curious behaviour is that the wavefunctions of the bound state and the continuum interfere either constructively or destructively, depending on their relative phase, and this is a function of the energy which changes across the line profile.

Inverse auto-ionization (also shown schematically in Fig. 2.20) is often an important astrophysical process. It is also known as dielectronic recombination. The first step is the capture by an excited ion of an electron into an excited state n^* to form a doubly excited state of the neutral atom. The recombination is stabilized by the return of the original excited electron to its ground state with the emission of a photon. This leaves a singly excited atom, the captured electron being still in the n^* state. The final step is the normal radiative decay of this singly excited atom. The importance of the effect is that the recombination cross-section has resonance bumps at energies corresponding to the excited states of the ion.

2.11 Hyperfine structure and isotope shift

The only property of the nucleus that has been considered so far is its charge, Z. This section shows how other properties of the nucleus affect the electronic energy levels and hence the spectrum. First, the energy levels depend on the mass and/or volume of the nucleus, so that analogous transitions in different isotopes

have slightly different frequencies. Secondly, any magnetic moment possessed by the nucleus by virtue of its intrinsic spin angular momentum interacts with the magnetic field produced by the angular momentum of the electrons. Thirdly, the nuclear electric quadrupole moment (if any) interacts with the gradient of the electronic electric field at the nucleus. There may even be interactions of higher order, but these are usually small enough to be ignored. The three effects specified are often of the same order of magnitude, and the term hyperfine structure (hfs), while generally applied to the magnetic dipole and electric quadrupole effects, is occasionally used to mean isotope shifts as well.

Let us consider first the nuclear magnetic moment $\boldsymbol{\mu}_I$ associated with a nuclear spin angular momentum $\boldsymbol{I}^*h$. I can take small integral or half-integral values according to whether the number of protons and neutrons (the mass number) is even or odd. $\boldsymbol{\mu}_I$ is proportional to $\boldsymbol{I}^*$, and the relation between them can be expressed in the same way as that for electron spin (equation 2.26)

$$\boldsymbol{\mu}_I = g_I \mu_N \boldsymbol{I}^* \tag{2.45}$$

μ_N is the *nuclear* magneton $e\hbar/2m_p$ where m_p is the proton rest mass. The nuclear g-factor defined by this equation is, like the Landé g-factor, of order unity, but it may be positive or negative. The nuclear magneton is smaller than the Bohr magneton by a factor m/m_p, or 1/1836, and μ_I is accordingly some three orders of magnitude smaller than μ_J. The energy level splitting is caused by the interaction of this magnetic moment with the magnetic field produced at the nucleus by the resultant angular momentum $\boldsymbol{J}^*h$ of the electrons. $\boldsymbol{B}_J$ is in the direction of $\boldsymbol{J}^*$ and must be of the same order of magnitude as the field B_L responsible for spin–orbit interaction (fine structure). The ratio of fine to hyperfine splitting is therefore determined essentially by the ratio of the magnetic moments μ_I/μ_J and is of order 10^{-3}. This means that hyperfine splittings are only a few hundredths of a wavenumber in the light elements, rising to a few wavenumbers in the heavy elements.

The coupling between $\boldsymbol{J}$ and $\boldsymbol{I}$ brings in another quantum number F determining the total angular momentum of the atom according to $\boldsymbol{F}=\boldsymbol{J}+\boldsymbol{I}$. Just as in the case of the L–S interaction, F can take the $2I+1$ values $J+I$, $J+I-1, \ldots, J-I$ (unless $J<I$, in which case there are $2J+1$ values from $I+J$ to $I-J$). The interaction energy is

$$\Delta E_{IJ} = \boldsymbol{\mu}_I \cdot \boldsymbol{B}_J = A\boldsymbol{I}^* \cdot \boldsymbol{J}^* \tag{2.46}$$

where

$$\boldsymbol{I}^* \cdot \boldsymbol{J}^* = \tfrac{1}{2}(F^{*2} - I^{*2} - J^{*2})$$

A level of given J is split into $2I+1$ (or $2J+1$) hyperfine levels whose separation is proportional to the hyperfine structure constant A. The analogy with fine structure is obvious: equation 2.46 can be obtained from equation 2.28 by substituting F, J, I for J, L, S, and the relative separations obey the Landé interval rule with F replacing J. Conventionally the same letter A is used for both the fine and hyperfine structure constants, despite the difference of some

three orders of magnitude between them. The actual value of the hyperfine A depends on g_I and on the electronic wave functions that determine the relation between B_J and J. A is at its largest for an unpaired s-electron, may be appreciable for a $p_{1/2}$-electron, and is usually negligibly small for $p_{3/2}$- and d-, f-, . . . , electrons. Figure 2.22 shows the hfs of the resonance lines of sodium, for which $I=\frac{3}{2}$. The selection rule for F needed to determine the line pattern is the same as for J: $\Delta F=0, \pm 1$, with $F=0 \rightarrow F=0$ forbidden.

Line splittings of the order 0.1–0.01 cm^{-1} are very difficult to measure accurately in the optical region, mainly because of the finite width of the lines themselves. Laser spectroscopy can overcome this by use of the 'Doppler-free' techniques described in Chapter 8. Alternatively, direct transitions between the hyperfine levels, which fall in the radiofrequency or microwave regions of the spectrum (remember that 0.1 $cm^{-1}=3$ GHz), can be accurately measured by magnetic resonance, level-crossing and double-resonance methods, as described in Chapter 9. We postpone the discussion of hyperfine structure in magnetic fields to that chapter.

The nuclear electric quadrupole interaction contributes to the hfs when the quadrupole moment Q and the gradient of the electric field produced at the nucleus by the electrons are both non-zero. The first condition requires $I \geqslant 1$; the second measures the departure of the electronic charge distribution from

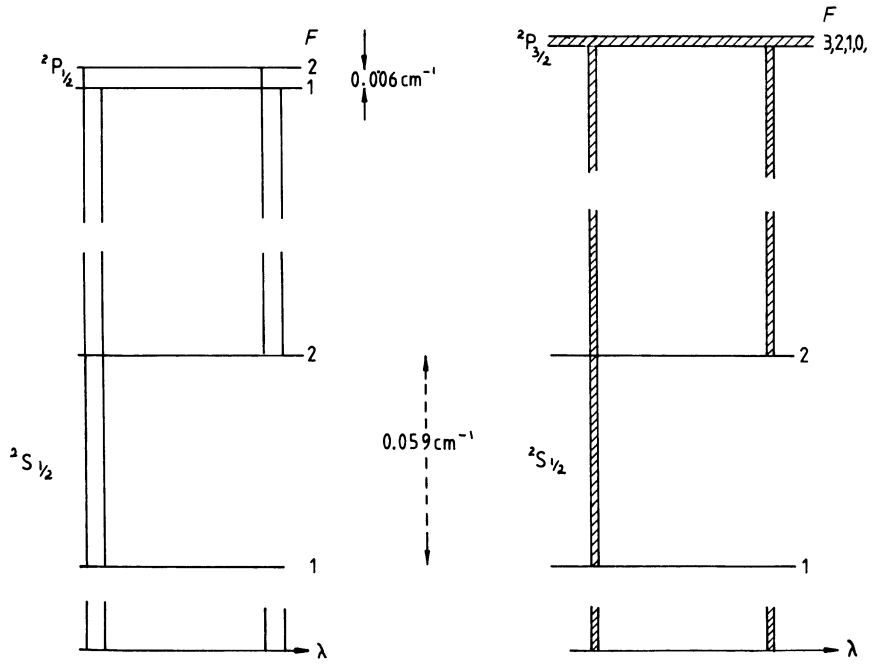

Figure 2.22 Hyperfine structure in the resonance lines of sodium ($I=\frac{3}{2}$). The splitting of the $^2P_{3/2}$ level is too small to show on the scale of the diagram.

spherical symmetry and requires $J \geqslant 1$. Assuming the asymmetry to be in the z-direction, the interaction energy is

$$\Delta E_Q = eQ\frac{\partial^2 V}{\partial z^2} f(F, I, J) \equiv Bf(F, I, J) \tag{2.47}$$

where $f(F, I, J)$ is a somewhat lengthy function of the three quantum numbers which may be found in [2, 12]. In free atoms departures from spherical symmetry are not large, so the quadrupole interaction constant B is generally smaller than the magnetic dipole constant A, and quadrupole effects usually show up only as small corrections on magnetic effects. In molecules and solids they are much larger because of the strong asymmetry of molecular and crystalline electric fields (section 9.3).

Isotope shift is the name given to the very small differences in energy found between corresponding electronic states of different isotopes. Since different levels are shifted by different amounts, the corresponding lines of two isotopes are slightly displaced from one another. In the lighter elements the effect is due to the difference in mass itself, which affects the Rydberg constant directly (equation 2.2). From equations 2.1 and 2.2, the shift $\Delta\sigma$ in a line of wavenumber σ for two isotopes differing in mass by ΔM is

$$\frac{\Delta\sigma}{\sigma} = \frac{\Delta E}{E} = \frac{\Delta\mu}{\mu} = \frac{m\Delta M}{M_1 M_2} \simeq \frac{m\Delta M}{M^2} \tag{2.48}$$

For hydrogen/deuterium this is about 10^{-3}. In all atoms except hydrogen and the hydrogen-like ions there may be in addition to this 'normal' shift a so-called specific mass effect that depends in a rather complicated way on the correlation of the electron motions and is very difficult to calculate reliably. Generally the normal and specific mass effects are of comparable size, but the specific effect may have the opposite sign. Because of the rapid decrease with increasing M, both effects can usually be ignored in the last third of the periodic table, but in the middle third the specific effect may be large enough to introduce significant uncertainties into the evaluation of the volume shift, which is the dominant effect for the heavier elements.

The volume, or field, effect is a result of the finite spread of the nucleus. Adding neutrons alters the size and sometimes the shape of the nuclear charge distribution and changes the electrostatic interaction with any electron charge distribution overlapping it. Only s-electrons, and to a lesser extent $p_{1/2}$-electrons, have an appreciable probability of being found near $r = 0$, and isotope shifts are confined to transitions in which the number of such electrons changes. Comparisons of theoretical with experimental absolute values are bedevilled by uncertainties in electron wave functions, screening constants and specific mass shifts, but measurements of relative shifts–the relative spacing between different isotopes of the same element for the same transition–have contributed much useful data to nuclear models, particularly since separated and short-lived

radioactive isotopes became available [11]. The shift between adjacent isotopes increases up the periodic table, rising to about a wavenumber with the heaviest elements. Since the radiofrequency methods cannot probe transitions between one isotope and another, they are inapplicable to isotope shift measurements, but the techniques of laser spectroscopy have virtually revolutionized the optical measurements. Furthermore, the possibility of using a tunable laser to excite selectively one particular isotope to a higher level, from which it can be readily ionized, has led to entirely new methods for isotope separation (section 8.5).

2.12 Ionic spectra

If one electron is removed from an atom the binding energy of the remaining electrons is increased because of the extra unbalanced nuclear charge. The ordinates of the energy level diagrams therefore have to be scaled up, transitions between levels require larger energy jumps, and the spectra are shifted to shorter wavelengths. In spectroscopic terminology the neutral atom is designated I, the singly ionized atom II, the doubly ionized atom III, and so on. A set of spectra in which the degree of ionization increases by one for each element along a row of the periodic table is known as an isoelectronic sequence because the number of electrons remains constant along it. The study of systematic changes along such a sequence is a valuable tool in spectral analysis, energy level calculations and the prediction of transitions in spectra that have not been observed or directly identified.

Energy levels for the hydrogen-like ions He^{2+}, Li^{3+}, Be^{4+}, . . . , can be calculated exactly, as noted in section 2.2. From equation 2.1 the hydrogen energy level diagram is simply scaled up by Z^2, where $Z = 2, 3, 4, \ldots$, and the wavelengths of the transitions are decreased by the same factor. The Lyman α-line $n = 2 \leftrightarrow n = 1$ at 122 nm in hydrogen occurs at 30.4 nm in He^{2+}, 13.5 nm in Li^{3+}, and so on. In all other isoelectronic sequences this simple relation is modified by the screening effects of the other electrons, but nevertheless the scaling factors along an iso-electronic sequence are approximately 4, 9, . . . , corresponding to z^2, where z is the 'unbalanced' nuclear charge, or the degree of ionization. As the interaction with the nucleus becomes relatively more important, the ions become increasingly hydrogen-like. They also become smaller, since $\langle r \rangle$ scales with $1/z$ (equation 2.3 and Table 2.3), and the magnetic effects (fine and hyperfine structure), which scale approximately with z^2 become relatively large.

Evidently, after a few stages of ionization one may expect the spectra to fall mainly in the far ultra-violet. Interest in and knowledge of ionic spectra have grown enormously since the advent of the high excitation laboratory sources described in Chapter 4, while at the same time observations from above the Earth's atmosphere have been made on stellar sources, especially, of course, the Sun. Lines below 200 nm from C, N, O and Si in the lower stages of ionization

are of particular astrophysical importance, as are lines from highly stripped ions of Fe and other heavier elements in the Sun's corona at wavelengths of a few tens of nm.

2.13 Inner shell and X-ray spectra

An electron in an inner shell sees a larger nuclear charge than does a valence electron. To a first approximation one would expect the same z^2 factor as in ionic spectra, where z is now the number of electrons in outer shells rather than the degree of ionization. The spectra therefore fall in the far ultra-violet. Not surprisingly, the central field model is not a good description of an atom in which electrons are excited from inner shells. Even with the inclusion of configuration mixing it may not be possible to calculate inner shell spectra with much reliability. The problem is that electron correlations may be important, and some form of many-body theory is then required.

If one goes right to the innermost shells the situation becomes simpler again, since here the potential is so dominated by the nucleus that the effects of the other electrons become relatively unimportant. Transitions involving these innermost shells constitute X-ray spectra. Since an electron in an $n=1$ shell sees almost the entire nuclear charge, energies scale approximately with Z^2. This relation was first recognized by Moseley, who showed that the frequencies of X-rays from different elements followed the law $\nu \propto Z^2$; a rather more accurate relation is $\nu \propto (Z-s)^2$, where s represents the screening effect of the electrons in the same or any lower shells. Since the same factor is applicable to an ion that has lost nearly all its electrons, X-ray spectra overlap in wavelength (in the region below 10 nm) the spectra of the highly stripped ions of the same elements. For the very light elements there is no hard-and-fast distinction between X-ray spectra and inner-shell excitation.

X-ray spectroscopy has grown up with its own notation: shells with $n=1, 2, 3, \ldots$, are designated K, L, M, ... For $n=1$ there is only one value of l, $l=0$, but for $n=2$ the two sub-shells $l=0$ and $l=1$ are denoted by L_1 and L_2, respectively. Similarly, there are three M sub-shells.

Another reason for the relative simplicity of X-ray spectra is that, since the inner shells are normally all full, transitions between them cannot take place until a vacancy is created. *Absorption* requires a photon of sufficient energy to remove an electron from the inner shell to an unoccupied outer shell, which is almost tantamount to removing it altogether from the atom. Thus the absorption spectrum consists of a series of steep steps, the K, L_1, etc., absorption edges. Normally these edges show some fine structure from the valence shells into which the electron may go; since one is normally concerned with solids, the structure gives information on the crystal fields rather than on the isolated atom, and is of the order of a few eV, compared with a total energy of some keV.

The characteristic X-ray *emission* spectra require first the ejection of an electron from an inner shell, either by a photon or, as in the normal X-ray tube,

by a fast electron. If the vacancy is formed in the K-shell, it can be filled by an electron from the L, M, . . . , shells with emission of sharp lines known as the K_α, K_β, . . . , lines. Similarly, a vacancy in the L-shell gives rise to the L_α, L_β, . . . , series. Evidently these series converge to limits coinciding in energy with the K, L, . . . , absorption edges. It was the Z^2 dependence of the K_α lines of successive elements in the periodic table that led Moseley to formulate his law in 1913, before a simple explanation in terms of shell structure was forthcoming.

Although the X-ray spectra of light elements fall in the same wavelength region as the spectra of highly stripped ions, in general the experimental methods and the applications of X-ray spectroscopy differ from those of optical and radiofrequency spectroscopy, and we shall not discuss X-rays further in this book.

References

FURTHER GENERAL READING (in alphabetical order)

1. Bransden, B.H. and Joachain, C.J. (1983) *Physics of Atoms and Molecules*, Longmans, London.
2. Kuhn, H.G. (1969) *Atomic Spectra*, Longmans, London.
3. Slater, J.C. (1968) *Quantum Theory of Matter*, McGraw-Hill, New York.
4. Woodgate, G.K. (1970) *Elementary Atomic Structure*, McGraw-Hill, New York.

RADIATIVE TRANSITIONS

5. Corney, A. (1977) *Atomic and Laser Spectroscopy*, Oxford University Press, Oxford.
6. Loudon, R. (1973) *Quantum Theory of Light*, Oxford University Press, Oxford.

MORE ADVANCED THEORY

*7. Condon, E.U. and Shortley, G. H. (1963) *The Theory of Atomic Spectra*, Cambridge University Press, Cambridge (reprint of 1935 book).
*8. Hibbert, A. (1975) Developments in atomic structure calculations, *Rep. Prog. Phys.*, **38**, 1217.
*9. Judd, B.R. (1985) Complex atomic spectra, *Rep. Prog. Phys.*, **48**, 907.
*10. Slater, J.C. (1960) *Quantum Theory of Atomic Structure*, Vols 1 & 2, McGraw-Hill, New York.

HYPERFINE STRUCTURE AND ISOTOPE SHIFT

11. King, W.H. (1984) *Isotope Shifts in Atomic Spectra*, Plenum Press, New York.
12. Kopfermann, H. (1958) *Nuclear Moments*, Academic Press, London.

For atomic energy levels see [1–3] of Chapter 1.

* Further reading, not cited in text.

THREE

A summary of molecular structure

3.1 Interatomic potentials

As two atoms approach one another the valence electrons on each begin to experience the attractive potential of the other nucleus, and a rearrangement of electron distribution takes place so as to minimize the total energy in this two-centre field. The effect is that of an attractive force between the two atoms, increasing with decreasing values of the internuclear distance R until R is small enough for repulsive forces to become important. These arise both from the internuclear potential $Z_1Z_2e^2/(4\pi\varepsilon_0R)$ and from the 'Pauli repulsion' due to the overlap of closed shells of electrons. The repulsive forces must eventually dominate the attraction, giving rise to a potential $V(R)$ of the form shown in Fig. 3.1: R_0 is the equilibrium internuclear distance and D_0 the dissociation energy of the molecule. For most simple molecules R_0 is a few ångströms (say 0.4–0.6 nm) and D_0 a few eV.

Figure 3.1 refers to the two atoms in their ground states. If one atom is in an excited state the baseline (the energy for $R=\infty$) is shifted up by the corresponding excitation energy, and a second interatomic potential curve is traced out as R is decreased. Because of the different electronic distribution in the excited atom this curve generally has a somewhat different shape, with different values of R_0 and D_0. In fact, several interatomic potentials are possible for a given electron configuration of each atom, depending on the coupling of the angular momenta of the electrons, just as different values of L, S and J give rise to different energy levels in a single atom.

The total energy of a diatomic molecule is a function not only of the electronic energy but also of the nuclear motion. Two isolated atoms have each three degrees of freedom, specified by motion along the three co-ordinate axes. When the two atoms are bound together, three of these degrees of freedom describe the motion of the centre of mass of the whole system. The other three are represented by relative motion: one by vibration along the internuclear axis and the other two by rotation about two axes perpendicular to the internuclear axis.

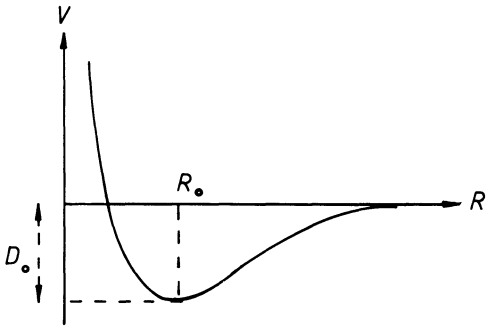

Figure 3.1 Interatomic potential. D_0 is the dissociation energy and R_0 the equilibrium internuclear separation.

Both the vibrational and rotational energies are quantized, as described in sections 3.5 and 3.6, and the wave functions describing the system are functions of the vibrational and rotational as well as the electronic quantum numbers.

3.2 The Born–Oppenheimer approximation

Schrödinger's equation for a molecule involves functions of the nuclear as well as the electronic co-ordinates, and the first simplification is to separate it into nuclear and electronic parts:

$$\psi = \psi_e(r_i, \boldsymbol{R}_N)\chi_n(\boldsymbol{R}_N)$$

where the $\boldsymbol{R}_N$ are the co-ordinates of the nuclei. It is then assumed that valid values of ψ_e can be found by taking constant values for $\boldsymbol{R}_N$, so that the Schrödinger equation can be separated into two equations, one involving ψ_e only and the other χ_n only. The solution of the ψ_e equation gives an energy E_e, which depends on the particular values of $\boldsymbol{R}_N$ chosen. This energy $E_e(\boldsymbol{R}_N)$ then forms part of the potential to be inserted into the second equation to give the total energy.

The justification for this, the Born–Oppenheimer approximation, is that the nuclei move much more slowly than the electrons, because of their much greater mass. In semi-classical terms, the time-scale associated with a Bohr orbit is of order $1/\nu$, the optical frequency, or $\sim 10^{-15}$ s, while that for an oscillation of the nuclei is of order 3×10^{-13} s, rotation being even slower.

A further simplification can be made by writing χ_n as the product of two functions, $\psi_v(R)$ representing changes in the relative positions of the nuclei (vibration along the internuclear axis in the case of a diatomic molecule) and $\psi_r(\theta, \varphi)$ representing rotation of the molecule. It is then possible to write

$$\begin{aligned} \psi &= \psi_e\psi_v\psi_n \\ E &= E_e + E_v + E_r \end{aligned} \tag{3.1}$$

The empirical justification for equation 3.1 is the condition $E_e \gg E_v \gg E_r$. It will be seen that vibrational quanta are usually a few thousand wavenumbers, or a few tenths of an eV, while rotational quanta are some two orders of magnitude smaller. Since the first excited states of simple atoms are usually several eV above their ground states, the separation is usually justified. However, the Born–Oppenheimer approximation breaks down when potential curves arising from different excited electron configurations are close to or even cross one another (section 3.8).

The electronic wave functions and energies are not, of course, those of the isolated atom(s). The different approaches to calculating them are described in the next section. If known with sufficient accuracy, they can be used to calculate $E_e(R)$, and this, together with the nuclear repulsion $Z_A Z_B e^2/(4\pi\varepsilon_0 R)$, forms the 'potential' curve of Fig. 3.1 which governs the nuclear motion. A semi-empirical potential is often used for this purpose, as described in section 3.4. Conversely, the rotational energy, which is determined to a first approximation by the moments of inertia of the molecule about the relevant axes of rotation, depends only on the equilibrium internuclear separations (and on the bond angles in the case of polyatomic molecules), and not on the shape of the interatomic potential. Rotation is discussed in section 3.5.

The rest of this chapter refers to diatomic molecules only. The complexities of vibration and rotation for polyatomic molecules are beyond the scope of this book, although a few remarks about them can be found in section 3.9.

3.3 Electronic energy of diatomic molecules

Assuming the validity of the Born–Oppenheimer approximation, we need to solve Schrödinger's equation for the electrons in the field of the two nuclei. The central field method is obviously inappropriate for two centres of force, but one can still use a self-consistent field approach: that is, one guesses at an initial wave function involving both atoms, A and B, uses the resultant electron distribution to calculate a potential distribution, solves Schrödinger's equation with this potential to find a new wave function, and so on. The internuclear separation R is a variable parameter in this process, so in principle the solution must be carried through for different values of R. The energy values thus obtained trace out a potential curve of the type shown in Fig. 3.1.

The availability of powerful computers has led to significant advances in quantum chemistry–that is, the calculation of molecular wave functions. Here we outline the two general approaches to the problem (see [4] for a full descriptive treatment and [12] for a review of methods of calculation). The *molecular orbital* approach treats each valence electron as belonging to both nuclei, usually by writing its wave function as a *linear combination* of the two *atomic orbitals* u_A and u_B (LCAO). There are necessarily two such combinations satisfying the requirement that the electron density be invariant against

exchange of electrons (compare the helium atom, section 2.6); one of these combinations increases the probability of finding the electron between the nuclei, where it can be thought of as gluing the nuclei together, while the other concentrates the electron density outside the internuclear region and has the reverse effect. These combinations are known as bonding and anti-bonding orbitals, respectively. The R-dependence of the energies calculated with them is of the form shown by the dotted curves in Fig. 3.2. Adding to these energies the internuclear repulsion $Z_A Z_B e^2/(4\pi\varepsilon_0 R)$ (broken curve) gives the resultant energies shown by the full curves. Evidently the bonding orbital, for which the energy passes through a minimum at some value R_0, is the one required for a stable molecule.

The simple LCAO treatment is not an entirely adequate description of the molecule, since it does not allow the electrons to return to their parent atoms as R is increased. To obtain a wave function with the right behaviour at large R it is necessary to add some fraction of the anti-bonding function back into the bonding function. The required fraction can be taken as a parameter to be adjusted to minimize the energy, on the general principle that the best wave function is the one that gives the lowest energy.

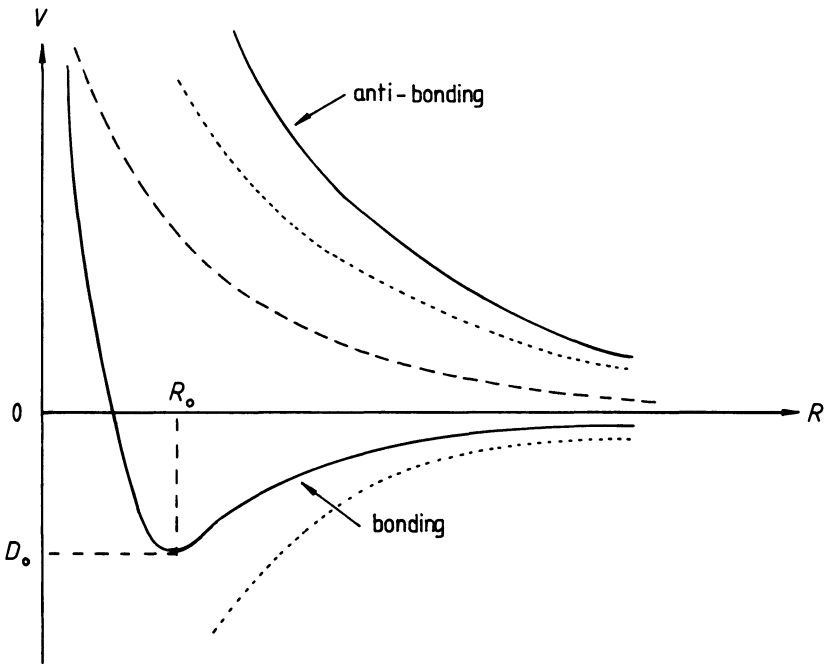

Figure 3.2 Electronic and nuclear contributions to interatomic potential. The dotted curves give the electronic energy for the bonding and anti-bonding orbitals, and the full curves show the effect of adding the internuclear repulsion (broken curve).

The second approach to the initial wave function, the *Heitler–London* method, starts with the separated atoms, but incorporates an interchange of electrons between them. Again, this results in one bonding state and one anti-bonding state. Again, it is not wholly satisfactory, but this time for the opposite reason: it does not allow sufficiently for the sharing of the electrons – i.e. the probability of finding both electrons near to one nucleus. The correction for this defect is to add some fraction of the separated ion configuration, with both electrons attached to one nucleus. In fact, this corrected Heitler–London function can be made identical with the corrected LCAO function by suitable choice of the mixing fractions, and the starting point is more a matter of computational convenience than of physical principle.

The filling of molecular orbitals is governed by the Pauli exclusion principle in the same way as that of atomic orbitals. Each orbital can contain two electrons, provided their spins are opposed. When the lowest energy (bonding) orbitals are full, any additional valence electrons must go into anti-bonding orbitals. For example, the two 1s-electrons from two hydrogen atoms both go into the bonding orbital $\psi = u_A(1s) + u_B(1s)$ with spins opposed. The two additional 1s-electrons in the combination He + He have to go into the anti-bonding orbital $\psi = u_A(1s) - u_B(1s)$, and the *net* bonding energy is effectively zero. p-electrons can form three distinct bonding orbitals, corresponding to the threefold degeneracy of the p-wave function ($l=1$, $m=0, \pm 1$). The oxygen atom has four 2p valence electrons, so in the O_2 molecule six of the eight valence electrons go into the three bonding orbitals and two into anti-bonding orbitals; the net score is two filled bonding orbitals, constituting a 'double bond'.

For a homonuclear molecule the equilibrium electron distribution is necessarily symmetric with respect to the two nuclei. The centroids of the electron and the nuclear charge distributions must coincide at the midpoint between the two nuclei, so the molecule has no permanent electric dipole moment. For a heteronuclear molecule the energy balance is generally favoured by a concentration of electron density near one nucleus. In NaCl, for example, the 3s-electron contributed by the sodium atom and the unpaired 3p-electron from the chlorine atom both go into a molecular orbital whose probability density is much greater near the chlorine atom than near the sodium atom. As a result the centroid of the electronic charge distribution is shifted towards the chlorine atom, and the molecule *does* possess a permanent dipole moment. Although to some extent such a molecule may be regarded as if it were formed of two well-defined ions, Na^+ and Cl^-, the electron transfer is by no means complete. There is an infinite gradation possible between pure covalent and pure ionic bonding, and the electric dipole moment of any molecule is a measure of its degree of ionic bonding. As shown later in this chapter, the appearance of vibrational and rotational spectra depend on the existence of this dipole moment.

If one or both of the atoms is in an excited state, further sets of bonding and anti-bonding orbitals can be constructed, giving potential curves similar to those of Fig. 3.2 but with the zero of energy ($R=\infty$) shifted up by the atomic

excitation energy. A few of the states of O_2 are shown in Fig. 3.3. It can be seen that several stable molecular states may result from each pair of atomic states: this is a result of the different ways in which their angular momenta may be coupled. A brief description of this coupling is necessary to understand the labelling of these states and the selection rules governing transitions between them.

The situation for a molecule is somewhat similar to that for an atom, in which a given electron configuration $(n_1 l_1, n_2 l_2, \ldots)$ gives rise to several different energy levels characterized by resultant angular momenta vectors $\boldsymbol{L}$, $\boldsymbol{S}$ and $\boldsymbol{J}$, but there is an important difference due to the presence of an electric field along the internuclear axis. The individual ls are quantized along this axis with components m_l, and it is the resultant $\sum m_l$ that is significant rather than L. This resultant is designated Λ, and states with $\Lambda = 0, 1, 2, \ldots$, are called Σ, Π, Δ, ... states – the Greek equivalents of the atomic S, P, D, ... The individual electron spins are rather weakly coupled to the internuclear axis and form a resultant S as in atoms; but for a rotationless molecule it is the axial component of S, designated Σ, that is coupled to Λ, forming a resultant Ω:

$$\Omega = |\Lambda + \Sigma| \tag{3.2}$$

A state with $\Sigma = 0$ and $\Lambda = 0$ is written as $^1\Sigma$; a state with $\Sigma = 1$ and $\Lambda = 1$ is written as $^3\Pi$, with the Ω-value as a subscript ($^3\Pi_2$, etc.), and so on.

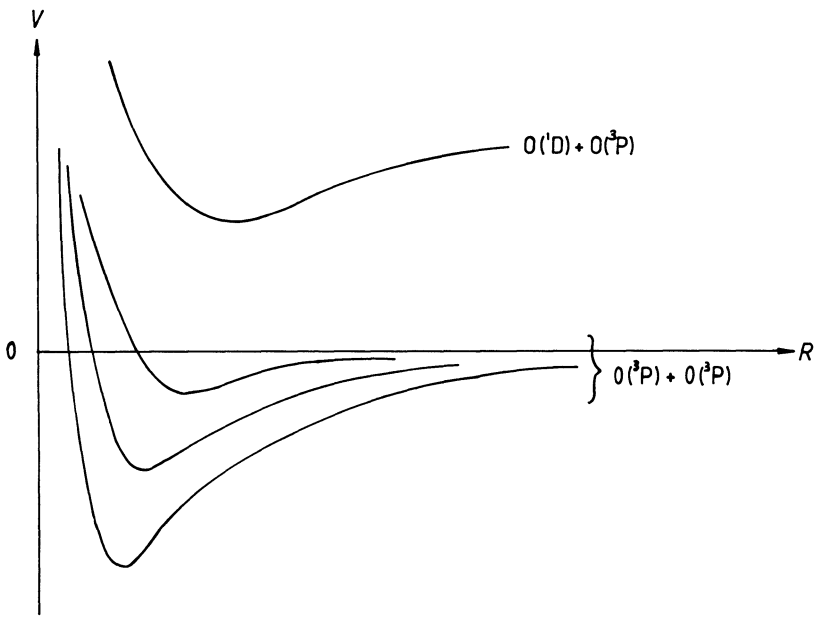

Figure 3.3 The lower electronic states of O_2. Only a few of the known states are shown.

There are two further complications for molecular states, both concerned with the symmetry of the electron orbital. The first affects homonuclear molecules only. The electron density $|\psi|^2$ in a homonuclear molecule must always be symmetric with respect to the midpoint between the nuclei, but the wave function itself can be either symmetric ($\psi \rightarrow \psi$)–gerade–or antisymmetric ($\psi \rightarrow -\psi$)–ungerade. The symmetry is shown by a subscript g or u. In the H_2 molecule, for example, the binding (stable) molecular states formed from the atoms H(1s) and H(2p) are

$$^1\Sigma_g \quad ^1\Sigma_u \quad ^1\Pi_u \quad ^3\Sigma_g \quad ^3\Sigma_u \quad ^3\Pi_u$$

Secondly, one has to take note of the symmetry with respect to reflection in any plane drawn through both nuclei. Again, the electron density $|\psi|^2$ is necessarily symmetric with respect to such a plane, in heteronuclear as well as homonuclear molecules, but ψ can go to either $+\psi$ or $-\psi$. Only one of these options is possible for any given Σ state, and a $+$ or a $-$ superscript is used to show which. The above Σ states for H_2 are all Σ^+ states. Both $+$ and $-$ symmetry occur in states with $\Lambda \neq 0$, but the $+$ and $-$ levels are degenerate until the molecule starts rotating. The origin of this degeneracy is essentially the same as that for atoms in an electric field (section 2.9): $+M$ and $-M$ have the same energy. The degeneracy is removed by the coupling between orbital and rotational angular momentum, resulting in a small splitting known as Λ-doubling.

In the ground states of many–especially homonuclear–molecules the electrons are paired off to make both Λ and S zero, giving $^1\Sigma$ states. Such molecules have no permanent magnetic dipole moment in their ground states. An important exception is O_2, for which the ground state is $^3\Sigma_g^-$, with spin 1. (This arises because the two anti-bonding electrons are in different anti-bonding orbitals and can therefore have parallel spins.) Molecular oxygen is thus paramagnetic. Molecules or molecular radicals with an odd number of valence electrons (e.g. NO, OH) form doublets ($S = \frac{1}{2}$) and are also paramagnetic.

3.4 Vibrational energy of diatomic molecules

The electronic energy curves of Figs 3.1–3.3 are the potential curves governing the nuclear motion. Classically, if the nuclei have zero kinetic energy they must be at rest at separation R_0, while if they have kinetic energy E_v they can oscillate between R_1 and R_2 (Fig. 3.4).

The principles on which calculations of the potential curve are based were described in the previous section. In practice the potential is frequently approximated by an empirical expression known as the Morse potential:

$$V = D_e(1 - e^{-\beta x})^2 \tag{3.3}$$

Here V is the energy referred to the minimum of the curve, D_e is the depth of the

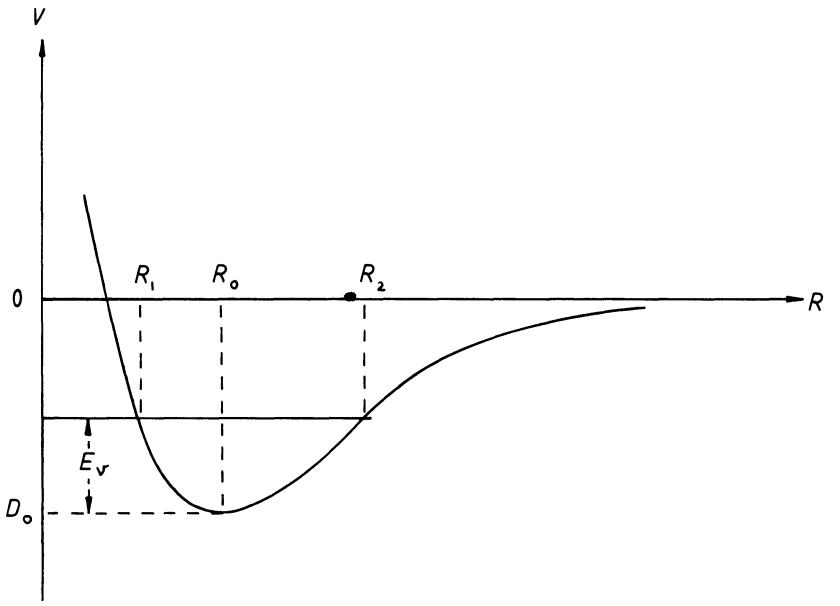

Figure 3.4 Vibrational energy of a diatomic molecule. R_1 and R_2 are the classical turning points for vibrational energy E_v.

well, $x(\equiv R-R_0)$ is the displacement of the nuclei from their equilibrium separation and β is a constant. For small displacements the potential can be expanded as

$$V=D_e\beta^2x^2\left(1-\frac{\beta x}{2}+\ldots\right) \tag{3.4}$$

In fact, any potential can be expanded in a similar way near its minimum, so we shall not lose much generality by sticking with the Morse potential in this section.

The leading term represents a parabola, so to first order the motion is simple harmonic with a restoring force given by

$$-\frac{\partial V}{\partial x}=-2D_e\beta^2x \tag{3.5}$$

The corresponding classical vibrational frequency is

$$\nu_0=\frac{1}{2\pi}\sqrt{\left(\frac{2D_e}{\mu}\right)}\beta \quad \text{with} \quad \mu=\frac{M_AM_B}{M_A+M_B} \tag{3.6}$$

and the classical vibrational energy is related to the amplitude of the vibrations and may have any arbitrary value. Quantum mechanically we have to solve

Schrödinger's equation for the variable x and the potential of equation 3.3. For the simple harmonic approximation the equation is

$$\left(-\frac{\hbar^2}{2\mu}\frac{\partial^2}{\partial x^2}+D_e\beta^2x^2\right)\psi_v=E_v\psi_v \tag{3.7}$$

This equation can be solved analytically, and the solutions may be found in any standard book on quantum mechanics. Physically well-behaved solutions are possible only for values of E given by

$$E_v=h\nu_0(v+\tfrac{1}{2}) \tag{3.8}$$

where ν_0 is the classical frequency given by equation 3.6 and the vibrational quantum number v takes integral values 0, 1, 2, ... The energy levels are evidently equally spaced; the first few are shown in Fig. 3.5.

For practical purposes the vibrational energy is usually expressed in wavenumbers and designated by G:

$$G(v)=\frac{E_v}{100ch}=\omega_e(v+\tfrac{1}{2})\quad \text{cm}^{-1} \tag{3.9}$$

where ω_e, the vibrational constant, is in units of cm^{-1}. ω_e depends on the

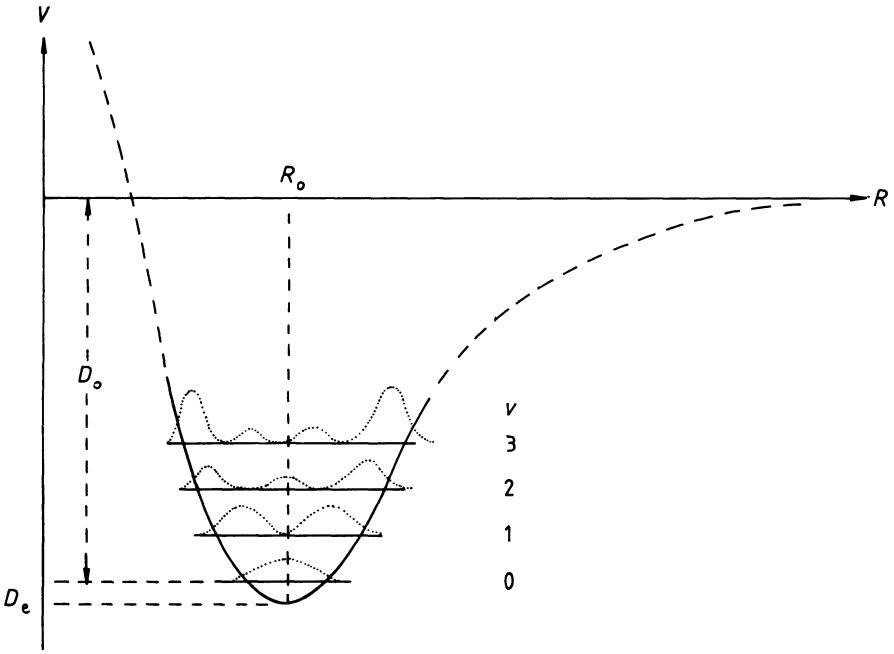

Figure 3.5 Vibrational energy levels and wave functions. The solid part of the potential curve is assumed to be parabolic. The dotted curves represent $|\psi_v(R)|^2$, the probability of a nuclear separation R.

particular electronic level considered as well as on the molecule, but is usually in the range 300–3000 cm^{-1}.

An important difference between the quantum and the classical oscillator is the existence in the former of the so-called zero-point energy, the vibrational energy $\frac{1}{2}\omega_e$ associated with the lowest level, $v=0$. The dissociation energy of the molecule, D_0, is measured from the lowest vibrational energy level and is therefore less than D_e, the depth of the well, by $\frac{1}{2}\omega_e$; this is typically around 0.1 eV, a not insignificant difference in a total of 2–3 eV. For comparison, kT at room temperature is 0.025 eV, or 200 cm^{-1}.

The general behaviour of the wave functions satisfying equation 3.7 is illustrated by the sketches of $|\psi_v|^2$ in Fig. 3.5. Remembering that these curves represent the relative probability of finding the nuclei at a given separation, it is evident that as v increases the nuclei are increasingly likely to be found near their extreme positions, the turning points of the classical motion.

When the full Morse potential (equation 3.4) is inserted in the Schrödinger equation, instead of just the first term in the expansion, the allowed energy levels are found to be

$$G(v)=\omega_e(v+\tfrac{1}{2})-x_e\omega_e(v+\tfrac{1}{2})^2 \quad cm^{-1} \tag{3.10}$$

where x_e, the anharmonicity constant, is small and positive (~ 0.01). The vibrational levels therefore crowd together as v increases, but they remain finite in number. They do *not* converge towards the dissociation limit as atomic energy levels converge towards an ionization limit. For large values of v, terms of yet higher order in v may be required to describe the energy correctly, since the Morse potential itself is not necessarily a good approximation for large displacements. The wave functions for the Morse potential are not, of course, the same as those of the simple harmonic oscillator, but their general form is still as shown in Fig. 3.5.

Values of ω_e and x_e for different electronic states of a very large number of molecules and molecular ions are given in tables in [7]. For updated and more comprehensive lists see [10].

3.5 Rotational energy of diatomic molecules

We start by treating the molecule as a rigid dumbbell, consisting of masses M_A and M_B joined by a rigid bar of length R_0, rotating about an axis through its centre of mass (Fig. 3.6). The moment of inertia I about this axis is given by $I=\mu R_0$, where μ is the reduced mass (equation 3.6). Classically, for an angular velocity ω the system has angular momentum $\boldsymbol{P}=I\boldsymbol{\omega}$ and energy $\frac{1}{2}I\omega^2=P^2/2I$. Quantum mechanically, one must solve the angular part of Schrödinger's equation for the nuclear co-ordinates to obtain the eigenvalues of the energy E_R, with corresponding wavefunctions ψ_r. The potential V in this equation is a function of R only, so for fixed R the angular equation is identical with that for

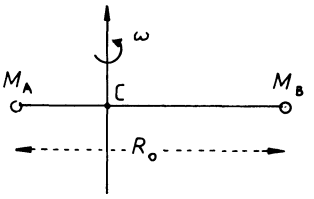

Figure 3.6 Rotation of a diatomic molecule with angular velocity about an axis through the centre of mass C.

the hydrogen atom. The solutions are the spherical harmonics $Y(\theta, \varphi)$ of equation 2.13, with the same restrictions on allowed values. In the molecular rotation case the angular momentum quantum number is designated J rather than l. The allowed values of angular momentum are then $\sqrt{[J(J+1)]}\hbar$, with $J = 0, 1, 2, \ldots$, and the z-component is $M\hbar$, with $M = J, J-1, \ldots, -J$. The energy eigenvalues are

$$E_r = \frac{1}{2I} J(J+1)\hbar^2 \tag{3.11}$$

which is identical with the classical relation between angular momentum and energy.

For practical purposes the rotational energy is expressed in wavenumbers and designated by the letter F:

$$F(J) = \frac{E_r}{100ch} = BJ(J+1) \quad \text{cm}^{-1} \tag{3.12}$$

where the rotational constant B is in cm^{-1} units and is given by

$$B = \frac{\hbar^2}{2I(100ch)} = \frac{\hbar}{4\pi I(100c)} = \frac{\hbar}{4\pi(100c)\mu R_0^2} \tag{3.13}$$

For two approximately equal nuclei of mass M, $\mu \simeq \frac{1}{2}M$, and we have seen that $R_0 \simeq 2 \times 10^{-10}$ m. Thus for $M \sim 40$ we have $B \sim 1\ \text{cm}^{-1}$. If one nucleus is much lighter than the other, $M_A \ll M_B$ say, then $\mu \simeq M_A$. A diatomic hydride such as HCl, NaH, etc., therefore has a particularly large B-value, in the range 10–20 cm^{-1}. For a given pair of nuclei the value of B varies from one electronic state to another because of the R_0-dependence.

The rotational energy levels given by equation 3.12 are 0, $2B$, $6B$, $12B$, . . . as shown in Fig. 3.7. The spacing between levels increases by $2B$ each time. However, the spacing is always much less than the room temperature energy of 200 cm^{-1}, in contrast with the spacing of vibrational levels.

We now consider corrections to this simple rigid rotator. First, the existence of vibration means that the internuclear separation is not fixed at R_0; even in the

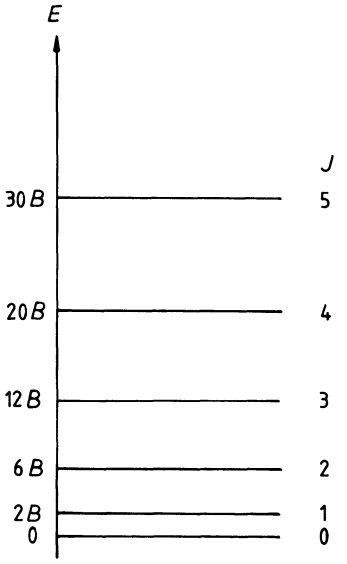

Figure 3.7 Rotational energy levels. B is the rotational constant (see text) and J the rotational quantum number.

lowest vibrational state there are small variations in R arising from the zero-point energy. From equation 3.13, B is determined by the expectation value of $1/R^2$, and B is therefore a function of v:

$$B_v = \frac{\hbar}{4\pi(100c)\mu}\int_0^\infty \psi_v^*(R)\frac{1}{R^2}\psi_v(R)\,dR \tag{3.14}$$

Evaluation of the integral shows that B_v can be represented by the expansion

$$B_v = B_0 - \alpha_e(v+\tfrac{1}{2}) + \gamma_e(v+\tfrac{1}{2})^2 + \ldots \quad \text{cm}^{-1} \tag{3.15}$$

where

$$\frac{\alpha_e}{B_0} = -\frac{6B_0}{\omega_e}$$

B_0 is never observed because of the zero-point energy. This dependence of rotational on vibrational energy levels represents a breakdown of the Born–Oppenheimer approximation.

The second correction arises from centrifugal distortion, which tends to increase R at high J-values. The corrected value of R is found by equating centrifugal force to the restoring force $D_e\beta^2(R-R_0)^2$ of equation 3.5; it leads to a correction term in the rotational energy proportional to $J^2(J+1)^2$:

$$F(J) = B_v J(J+1) - D_v J^2(J+1)^2 \quad \text{cm}^{-1} \tag{3.16}$$

where

$$D_v = \frac{4B_0^3}{\omega_e^2}$$

With $B_v \sim 1\ \text{cm}^{-1}$ and $\omega_e \sim 1000\ \text{cm}^{-1}$, the correction factor for $J \sim 30$ is a few tenths of a percent. Experimental values for the rotational constants are given in [7, 10].

So far in this section the angular momentum $J\hbar$ has been assumed to come entirely from nuclear rotation, but if the electronic angular momentum is not zero, J must be interpreted as the resultant of all three contributions–rotational, spin and orbital. There are several different ways of coupling these angular momenta, according to the relative strengths of the interactions. The two most important are known as Hund's cases (a) and (b). In case (a) the rotation does not disturb the coupling of the electronic motion to the internuclear axis, so Ω in equation 3.2 remains well defined and is coupled to the nuclear rotation angular momentum vector N to give

$$J = \Omega + N \tag{3.17}$$

In case (b) the spin S is not coupled to the internuclear axis, so Ω is not defined. Λ and N are coupled to form a resultant K, and this in turn is coupled to the spin S to give the final resultant J:

$$J = K + S = (\Lambda + N) + S \tag{3.18}$$

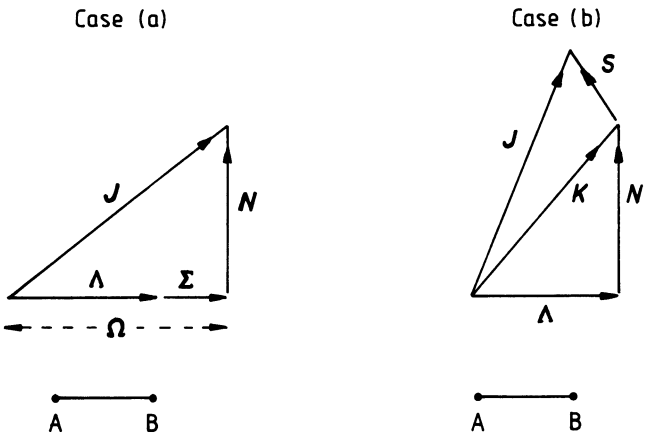

Figure 3.8 Coupling of angular momenta for Hund's cases (a) and (b). AB is the direction of the internuclear axis. Λ is the component of orbital angular momentum along AB, S and Σ are the spin and its component along AB, respectively, N is the angular momentum from nuclear rotation, and J is the final resultant.

Figure 3.8 illustrates these two cases schematically. It can be shown that the rotational energies are:

$$\left.\begin{array}{ll} \text{case (a)} & F(J)=B[J(J+1)-\Omega^2] \\ \text{case (b)} & F(J)=BK(K+1)\pm\delta \end{array}\right\} \tag{3.19}$$

where δ represents the spin splitting.

In practice the coupling may be intermediate between these cases (compare *L–S* and *j–j* coupling in section 2.7), or may make a transition from case (a) to case (b) with increasing rotation, or may conform to a scheme different from either of these two. For further details see [1, 2, 7, 11].

3.6 Spectra of diatomic molecules

Because of the very different orders of magnitude of the spacings of the electronic, vibrational and rotational energy levels, molecular transitions occur in three distinct regions of the spectrum. The pure rotational spectrum, transitions between rotational levels of the same vibrational and electronic state, is found in the microwave region – say 0.1 mm (100 μm) to 30 cm. The vibration–rotation spectrum, transitions between different vibrational levels of the same electronic state, with a fine structure composed of the rotational transitions, falls in the infra-red – say 100–1 μm. Transitions between different electronic states occur, as for atomic spectra, in the near infra-red, visible and ultra-violet regions. Molecular transitions are traditionally written with the upper state first, which is the opposite of the atomic convention. A single prime designates the upper state, and a double prime the lower state.

Starting with the simplest spectrum, radiative transitions between rotational levels can take place only if the molecule possesses a permanent electric dipole moment. This requirement is easy to understand classically in terms of radiation from a rotating dipole, and we look at the wave-mechanical representation in the next section, where the origin of the selection rule $\Delta J=\pm 1$ is also explained. From the considerations of section 3.3, homonuclear molecules do not have electric dipole moments, so do not have pure rotational spectra. For heteronuclear molecules the transitions allowed by the J selection rule gives rise to frequencies $2B$, $4B$, $6B$, . . . (neglecting the centrifugal correction), as is evident from Fig. 3.7 and equation 3.12. The pure rotational spectrum therefore consists of almost equally spaced lines separated by $2B$.

The infra-red spectrum due to transitions between vibrational levels similarly exists only for heteronuclear molecules, the requirement in this case being a dipole moment that varies with the internuclear distance. Again, the requirement for an oscillating dipole is an obvious one classically, and its wave-mechanical form is given in the next section. The selection rule for a pure harmonic oscillator is $\Delta v=\pm 1$, but for an anharmonic oscillator $\Delta v=\pm 2, 3, \ldots$ are also allowed as less probable transitions. The infra-red spectrum of a heteronuclear molecule therefore consists of a fundamental at wavenumber

ω_e (equation 3.9) with weaker overtones close to $2\omega_e$, $3\omega_e$, ... The multiples are not exact because of the anharmonicity correction (equation 3.10). The rotational structure turns each vibrational transition into a band, or a group of lines spaced at approximately equal intervals $2B$ on either side of the band origin σ_0. Figure 3.9 shows the first few rotational levels for each of two vibrational states v' and v'', with the corresponding spectrum underneath. Lines for which $J' = J'' + 1$ fall on the high-frequency side of σ_0 and are known as the R-branch, while the low-frequency side, $J' = J'' - 1$, is the P-branch. There is a

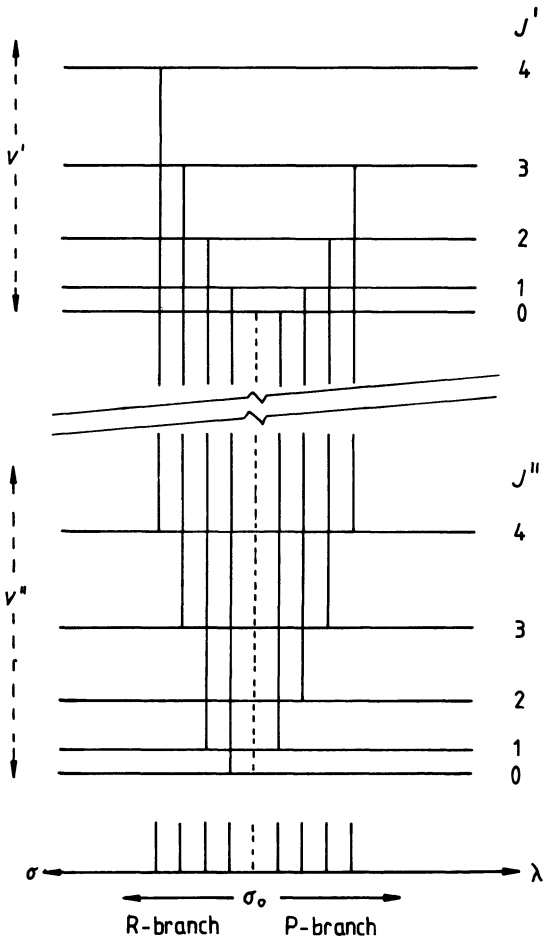

Figure 3.9 Vibration–rotation band. The diagram is drawn for $B'' \simeq B'$ so that the lines are almost equally spaced at $2B''$ (or $2B'$). The transition $J' = 0 \leftrightarrow J'' = 0$ (shown broken) is forbidden.

gap in the position of σ_0, which corresponds to the forbidden transition $J=0 \leftrightarrow J=0$. Again disregarding the D-correction, the wavenumbers of the lines are

$$\sigma = \sigma_0 + B'J'(J'+1) - B''J''(J''+1) \tag{3.20}$$

which for $J'=J''+1$ and $J'=J''-1$, respectively, becomes

$$\sigma_R(J'') = \sigma_0 + 2B' + (3B'-B'')J'' + (B'-B'')J''^2 \tag{3.21a}$$

$$\sigma_P(J'') = \sigma_0 + (B'+B'')J'' + (B'-B'')J''^2 \tag{3.21b}$$

In general $B' \neq B''$, because as we have seen the moment of inertia varies with the vibrational state, but if the difference is small these expressions simplify to

$$\left.\begin{aligned} \sigma_R(J'') &= \sigma_0 + 2B + 2BJ'' \\ \sigma_P(J'') &= \sigma_0 - 2BJ'' \end{aligned}\right\} \tag{3.22}$$

The electronic spectrum takes the form of systems of bands. Figure 3.10 shows schematically part of one such system, a few bands of the transition between two electronic states A and X. Each pair of vibrational levels (v', v'') may give rise to a band, and each band is composed of transitions between different rotational levels J'–J'', only a few of which are shown in the figure. The electronic spectrum appears even if the molecule does *not* have a permanent dipole moment, for the same reason that atomic transitions occur: the change in electron distribution between the two states provides the necessary 'transition moment', as described a little more fully in the next section. The vibrational and rotational energies of homonuclear molecules can therefore be investigated in the electronic transitions.

The selection rules governing the electronic transitions are obtained from the same type of symmetry consideration as for isolated atoms. For electric dipole transitions the condition that $\int \psi_e^* e r \psi_e \, d\tau$ does not vanish becomes, for the equivalent of L–S coupling,

$$\left.\begin{aligned} &\Delta\Lambda = 0 \quad \text{for light polarized along the internuclear axis} \\ &\Delta\Lambda = \pm 1 \quad \text{for light polarized perpendicular to the axis} \end{aligned}\right\} \tag{3.23}$$

with the additional restriction $\Delta\Sigma = 0$. For the equivalent of j–j coupling, where Λ and Σ are not individually defined, the corresponding rule is $\Delta\Omega = 0, \pm 1$. The additional restrictions applying to the $(+, -)$ designations of Σ states and the (g, u) designations of homonuclear molecules are

$$\begin{aligned} &+ \leftrightarrow + \quad \text{and} \quad - \leftrightarrow - \quad \text{but} \quad + \nleftrightarrow - \\ &\mathrm{u} \leftrightarrow \mathrm{g} \quad \text{but} \quad \mathrm{u} \nleftrightarrow \mathrm{u} \quad \text{and} \quad \mathrm{g} \nleftrightarrow \mathrm{g} \end{aligned} \tag{3.24}$$

We now look more closely at the band structure of the allowed electronic transition $A - X$ in Fig. 3.10. Neglecting the anharmonic terms in the vibrational

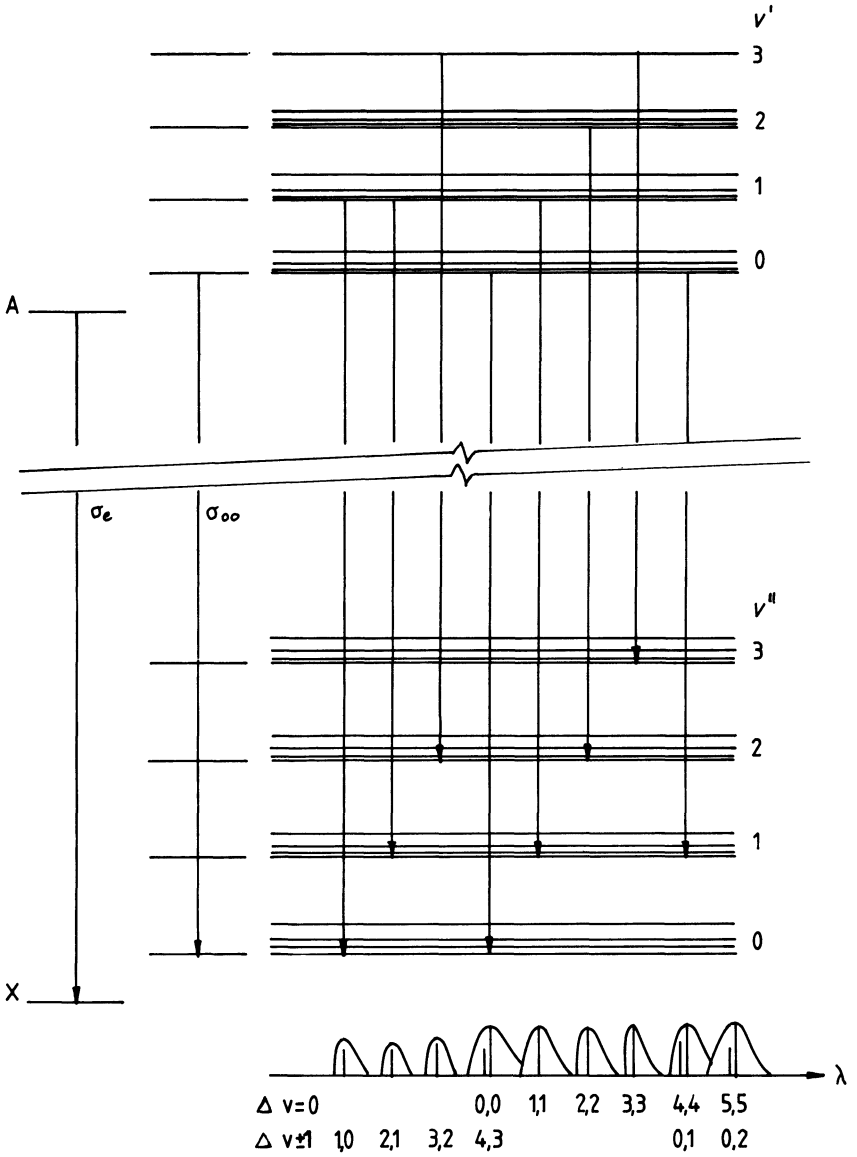

Figure 3.10 Electronic transition giving rise to a band spectrum. Only the first few rotational levels of each vibrational level are shown, and their spacing is greatly exaggerated. The vertical arrows, and the corresponding vertical lines in the schematic spectrum, show the band origins. The bands are formed by the rotational structure; they may extend a considerable distance from the origin and may overlap one another. The diagram is drawn for $\omega' < \omega''$.

energy, the wavenumbers of the bands are given by equation 3.10:

$$\begin{aligned}\sigma(v', v'') &= \sigma_e + \omega'(v' + \tfrac{1}{2}) - \omega''(v'' + \tfrac{1}{2}) \\ &= \sigma_{00} + \omega' v' - \omega'' v''\end{aligned} \tag{3.25}$$

where σ_e is the separation of the potential minima of the A and X states and σ_{00} is the wavenumber of the (0, 0) band. For electronic transitions there is no selection rule for Δv, so the entire band system may extend over several units of ω, or 100 nm or more. If $\omega' \simeq \omega''$, bands with the same Δv fall almost on top of one another. However, the number of strong bands is determined by the levels that are initially populated and by the transition probabilities from those levels. The latter are governed by the Franck–Condon principle, which is illustrated in Fig. 3.11. The potential curves for the two states A and X are shown with their first few vibrational levels. The wave functions sketched in Fig. 3.5 show that the

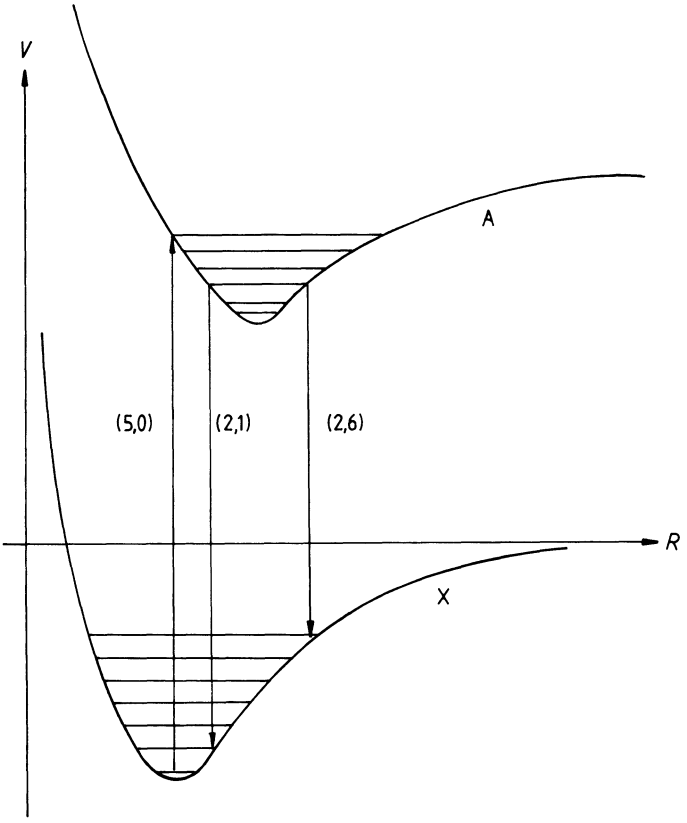

Figure 3.11 Illustration of Franck–Condon principle. The most probable absorption from $v'' = 0$ is to $v' = 5$. The most probable emissions from $v' = 2$ are to $v'' = 1$ and 6.

most probable values of R for any given v are near the classical 'turning points' at the ends of the horizontal lines, except for $v=0$ which has one central maximum. By the Born–Oppenheimer approximation the nuclei are to be taken as essentially fixed during an electronic transition. It follows that the most probable vibrational transitions are those between levels whose probability maxima occur at about the same R. If, as is frequently the case, only the lowest one or two vibrational levels are populated, the strongest transitions in absorption will be (for the system shown in Fig. 3.11) around (5, 0) and in emission around (0, 4), as indicated by the vertical lines.

Finally, we come to the rotational structure of a single band. Relative to the band origin σ_0 we have, as before (equation 3.20),

$$\sigma=\sigma_0+B'J'(J'+1)-B''J''(J''+1)$$

For electronic transitions the selection rules allow $\Delta J=0$ as well as $\Delta J=\pm 1$, except for transitions between two Σ states. In addition to the P- and R-branches given by equations 3.21 there is now a Q-branch, $\Delta J=0$:

$$\sigma_Q=\sigma_0+(B'-B'')J''+(B'-B'')J''^2 \tag{3.21c}$$

The Q-branch overlaps either the R-branch or the P-branch according to whether $B'>B''$ or $B'<B''$. Another distinction from the infra-red band structure is that the Bs of different electronic states are often very different, so that the J^2 term in the set of equations 3.21 is much more significant. From these equations it can be seen that if $B'>B''$ the terms in J^2 and J have opposite signs for the P-branch, which must therefore double back on itself to form a head. If $B'<B''$, the same thing occurs for the R-branch instead. In the first case the turning point is at the long-wavelength side of the band, which is said to be degraded to the violet; in the second case it is at the other end, and the band is degraded to the red. Photographs of both types of band, together with wavelengths of band heads for many molecules are given in [13].

In states other than Σ, the coupling of rotational with electronic angular momentum, which gives rise to the level splitting mentioned in the last section, also results in additional rotational branches [2, 7, 11]. A band may therefore consist of several hundred lines, very closely packed in the region of the band head but extending over several hundred cm^{-1}. A further point to be considered in the case of homonuclear molecules is the symmetry under *nuclear* interchange: we return to this point in section 3.8, where it is seen that its effect is to weaken, or eliminate, every alternate rotational line.

Rotational and vibration–rotational transitions in homonuclear molecules, which cannot occur by absorption or emission of radiation, can be investigated by the inelastic scattering of radiation, better known as Raman spectroscopy. The requirement for a Raman-active molecule is a polarizability that changes with rotation and vibration, and this requirement is met by homonuclear as well as heteronuclear molecules. Raman spectroscopy is discussed in section 9.10.

3.7 Transition probabilities for diatomic molecules

From equations 2.32 and 2.33, the probability of an atomic electronic transition is proportional to the square of the electric dipole transition moment:

$$P_{if} \propto |\langle u_f^* | e\boldsymbol{r} | u_i \rangle|^2$$

For a diatomic molecule the us must be replaced by the total wavefunctions ψ, and it is convenient to label the transition moment as R_{if}: i.e.

$$R_{if} = \langle \psi_f^* | e\boldsymbol{r} | \psi_i \rangle \quad \text{with} \quad P_{if} \propto |R_{if}|^2 \tag{3.26}$$

Using the Born–Oppenheimer approximation and equation 3.1,

$$\psi = \psi_e(\boldsymbol{r}, R)\, \psi_v(R) \psi_r(\theta, \varphi)$$

together with the prime and double-prime notation (it is immaterial for this purpose whether the transition is upward or downward),

$$R_{if} = \int \psi_e'^* \psi_v'^* e\boldsymbol{r} \psi_e'' \psi_v'' \, d\boldsymbol{r} \, dR \int \psi_r^{*\prime} \psi_r'' \sin\theta \, d\theta \, d\varphi \tag{3.27}$$

Separation of the integral in this way is possible because the second integral is a function of the nuclear angular co-ordinates θ and φ only.

For *pure rotational* transitions $\psi_e' = \psi_e''$ and $\psi_v' = \psi_v''$. The quantity

$$R_e(R) \equiv \int \psi_e^* e\boldsymbol{r} \psi_e \, d\boldsymbol{r} \tag{3.28}$$

is the permanent electric dipole moment of the molecule. It does not depend strongly on R, so the first integral in equation 3.27 becomes

$$\int \psi_v^* R_e \psi_v \, dR \simeq R_e \int \psi_v^* \psi_v \, dR = R_e$$

(for normalized vibrational wave functions). This confirms mathematically what we already know physically: pure rotational transitions require a permanent electric dipole moment. If this exists, the selection rules are obtained from the well-known properties of the spherical harmonics in the second integral of equation 3.27. By direct analogy with the atomic case (equation 2.34)

$$J'' = J' \pm 1 \quad \text{and} \quad M'' = M', M' \pm 1 \tag{3.29}$$

For *vibrational* transitions within the same electronic state equation 3.27 becomes

$$R_{if} = \int \psi_v'^* R_e(R) \psi_v'' \, dR \int \psi_r'^* \psi_r'' \sin\theta \, d\theta \, d\varphi \tag{3.30}$$

with $R_e(R)$ still defined by equation 3.28. The dependence of R_e on R can be expressed by expanding it as a power series:

$$R_e(R) = R_e(R_0) + \frac{\partial R_e}{\partial R}(R - R_0) + \cdots$$

When this is substituted in equation 3.30 the first term vanishes because of the orthogonality properties of the vibrational wave functions, and the second term

gives (omitting the angular factor)

$$R_{if} = \frac{\partial R_e}{\partial R} \int \psi_v'^* (R - R_0) \psi_v'' \, dR \tag{3.31}$$

Evidently this is zero unless R_e varies with R. If the ψ_v are harmonic oscillator functions, the integral is zero except for $\Delta v = \pm 1$. Rotation–vibration spectra therefore require $\partial R_e/\partial R \neq 0$, with the selection rules of equation 3.29 and, less rigidly, $\Delta v = \pm 1$.

Finally, for *electronic* transitions equation 3.27 becomes

$$R_{if} = \int \psi_v'^* R_e^{if}(R) \psi_v'' \, dR \int \psi_r'^* \psi_r'' \sin\theta \, d\theta \, d\varphi \tag{3.32}$$

The difference from equation 3.30 is that R_e^{if} is not a permanent dipole moment; it is a *transition moment* defined by

$$R_e^{if}(R) = \int \psi_e' e r \psi_e'' \, d\boldsymbol{r} \tag{3.33}$$

Again, it does not usually change very rapidly with R, so to a first approximation it may be replaced by an average value $\bar{R}_e$ and taken outside the integral. To the extent that this approximation is valid, R_{if} becomes the product of three integrals, each involving electronic, vibrational or rotational functions only:

$$R_{if} = \bar{R}_e^{if} \int \psi_v'^* \psi_v'' \, dR \int \psi_r'^* \psi_r \sin\theta \, d\theta \, d\varphi \tag{3.34}$$

The second term in this expression is known as the overlap integral, and it is large only when ψ_v' and ψ_v'' have maxima at approximately the same values of R. This is the wave-mechanical basis of the Franck–Condon principle. The last term is superficially the same as in equations 3.27 and 3.30, but there is, or may be, one difference. In treating ψ_r as representing pure nuclear rotation, depending on the quantum numbers J and M only, no account was taken of any electronic orbital angular momentum. In the ground electronic state, which is where pure rotation and vibration–rotation spectra are observed, this neglect is usually justified because the ground state is so often a $^1\Sigma$ state. However, if Λ is not zero in one or other of the electronic states, then Λ or Ω, according to the appropriate coupling scheme, enter into the rotational wave function and modify the selection rules and the value of the integral. In particular, $\Delta J = 0$ is allowed, as noted in the previous section. Further discussion of line and band strengths in diatomic molecular spectra can be found in section 11.10.

3.8 Some further points on spectra of diatomic molecules

3.8.1 MISSING ROTATIONAL LINES AND ALTERNATION OF INTENSITIES

Homonuclear molecules with zero nuclear spin are found to have alternate rotational lines missing in every band. If the nuclear spin is not zero, the rotational lines alternate between weak and strong in intensity. These phenomena are a consequence of the requirement that $|\psi^2|$ must be invariant against

exchange of identical nuclei – i.e. either $\psi \rightarrow \psi$ or $\psi \rightarrow -\psi$, the functions being labelled symmetric (ψ_s) or antisymmetric (ψ_a), respectively. It is shown below how the symmetry is determined. The effect on the spectrum is due to the selection rule

$$\psi_s \nleftrightarrow \psi_a \tag{3.35}$$

which is rigorous for all types of transition if the nuclear spin is zero or if it is coupled only weakly to the spatial co-ordinates.

The symmetry of ψ is determined as follows. From equation 3.1,

$$\psi = \psi_e(r, R)\, \psi_v(R)\, \psi_r(\theta, \varphi)$$

Interchanging the nuclei is tantamount to $\theta \rightarrow \theta + \pi$. This leaves $\psi_v(R)$ unaffected; in ψ_r the even terms in the spherical harmonics ($J = 0, 2, 4, \ldots$) are unchanged, while the odd terms ($J = 1, 3, \ldots$) change sign. For the electronic function, interchange of nuclei is equivalent to the two operations of inversion with respect to the midpoint between the nuclei – symmetry specified by g or u – and reflection in a plane through both nuclei – symmetry specified by $+$ or $-$ (section 3.3). ψ_e is therefore symmetric for the combinations ($+$, g) and ($-$, u), leading to symmetric ψ for even J and antisymmetric ψ for odd J. Conversely, ψ_e is antisymmetric for the combinations ($+$, u) and ($-$, g), and in this case ψ is symmetric for odd J and antisymmetric for even J.

Equation 3.24 specified as selection rules for electric dipole transitions $+ \leftrightarrow +$, $- \leftrightarrow -$ and g$\leftrightarrow$u. ψ'_e and ψ''_e must therefore have opposite symmetry; consequently, in one state even J levels are symmetric, and in the other odd J levels are symmetric. Equation 3.35 is thus satisfied for $\Delta J = \pm 1$, the P- and R-branches. At first sight it appears that $\Delta J = 0$ must be forbidden by equation 3.35, so that there would be no Q-branch. However, states with $\Lambda \neq 0$ (i.e. $\Pi, \Delta, \ldots$ states) are actually all double, either degenerate or nearly so, as explained in section 3.3. Since one of the pair is always $+$ and the other $-$, one of them must have the correct symmetry for $\Delta J = \pm 1$ and the other for $\Delta J = 0$. The Q-branch therefore exists for transitions other than Σ–Σ.

It remains to link the symmetry of ψ with the alternation of intensities of the rotational lines. Evidently, if symmetric and antisymmetric states are found with equal probability no additional restrictions are imposed on homonuclear molecules. However, quantum statistics requires that the total wavefunction for two identical particles be symmetric against interchange for bosons and antisymmetric for fermions. This rule was used in section 2.6 to explain the singlet–triplet structure in atoms with two valence electrons, the requirement there being antisymmetry of the total function because electrons are fermions. Nuclei are bosons or fermions according to whether their spin is integral or half-integral. For the particular case $I = 0$ the nuclei are bosons, and *only* the symmetric state ψ_s exists. Hence, all the even rotational levels will be missing from one state of an electronic transition and all the odd levels from the other – that is, alternate rotational lines will be missing.

For $I \neq 0$ one has to consider also the symmetry of the nuclear spin function. The function ψ that we have been calling the total molecular wave function involves only the *spatial* co-ordinates of the electrons and the nuclei, $\boldsymbol{r}$ and R, and we can take account of nuclear spin by multiplying it by an appropriate spin function, just as was done for the electron wave function in an atom (section 2.6). In the particular case of $I = \frac{1}{2}$ the symmetry of the spin functions is determined in exactly the same way as for electrons: there are three symmetric combinations and one antisymmetric, as shown in Table 2.4. Moreover, since such nuclei are fermions, the allowed combinations with the spatial functions are exactly as shown in Table 2.4, and therefore ψ_a has three times the statistical weight of ψ_s. The rotational lines of such nuclei show a 3:1 alternation in intensity.

For any arbitrary spin I there are $2I + 1$ spin states for each nucleus, and hence $(2I + 1)^2$ combinations for the two nuclei. It can be shown that $(I + 1)(2I + 1)$ of these are symmetric and $I(2I + 1)$ are antisymmetric, leading to an intensity ratio for alternate lines of $I/(I + 1)$.

3.8.2 CONTINUOUS SPECTRA

Continuous molecular spectra, arising from transitions between a discrete and a continuous state, or between two continuous states, occur in both absorption and emission. In atomic spectra a continuum is always associated with ionization ($A \rightarrow A^+ + e^-$); in molecular spectra it can also be associated with ionization ($AB \rightarrow AB^+ + e^-$), but it is more frequently due to dissociation ($AB \rightarrow A + B$). In this case the upper, the lower or both states may be continuous. Figure 3.12 shows how continuous absorption arises in the most common cases in which the upper state is either (a) unstable or (b) stable but with its minimum sufficiently displaced that transitions to high v' levels are probable by the Franck–Condon principle. In the latter case the continuum may join on to the discrete band spectrum. Some molecules, such as the inert gases, show continuous absorption from an unstable ground state to a bound upper state, as shown in Fig. 3.12(c). The ground state can be regarded as a temporary molecule formed by the collision of two atoms weakly attracted to one another by van der Waals forces, and absorption may occur at any value of R as one atom passes the other. Thus, absorption to a large spread of v' values is possible, and the absorption to any given v' is smeared out by the spread in kinetic energies of the colliding atoms. It can be seen from the figure that this type of continuum occurs to the long-wavelength side of the corresponding atomic absorption line.

It is clear from this last case that the distinction between molecular spectra and pressure – i.e. collisionally – broadened atomic spectra may become rather blurred. We have only to consider the upper state also to be a weakly bound van der Waals state (Fig. 3.12(d)) to see that there is actually *no* distinction between a temporary molecule and a pair of colliding atoms. Radiative transitions between the two states are closely related to the atomic transitions

$$A + B + h\nu \rightarrow A^* + B \quad \text{or} \quad A + B^*$$

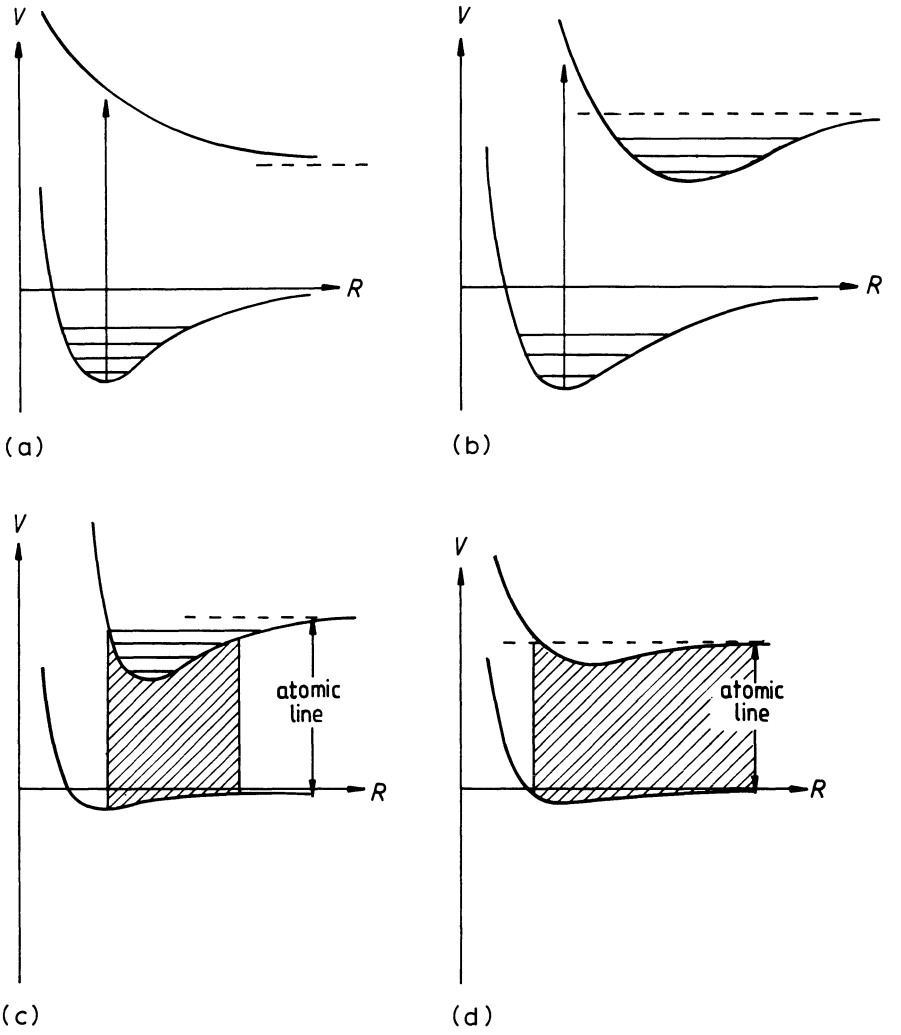

Figure 3.12 Continuous absorption:
(a) from a bound lower state to an unstable upper state;
(b) from a bound lower state to a bound upper state, above the dissociation limit;
(c) from an unbound lower state to a bound upper state;
(d) from an unbound lower state to an unbound upper state.
In (c) and (d) the continuum is associated with an atomic transition in the isolated atom.

As the Franck–Condon principle still holds, the transitions are represented by vertical lines on the diagram, and the atomic line will extend to longer or shorter wavelengths according to whether the upper or the lower curve has the deeper minimum. We return to this subject in section 10.5.

Continuous emission spectra can in principle be illustrated by reversing the arrows in Fig. 3.12. However, the cases shown in (a) and (b) require further consideration. The classical picture is of two atoms, one or both of them excited, approaching one another from infinity and radiating their excitation energy so as to drop to the lower (stable) state before they bounce apart again. Typically, the time taken for one oscillation (i.e. vibration) is about 10^{-13}s, as against 10^{-7}–10^{-8} s for a radiative lifetime, so radiation is an unlikely process. The molecule is much more likely to be stabilized by collision with a third body to carry away the excess energy.

This objection does not apply to the continuous emission corresponding to Figs 3.12(c) and (d). The quasi-continuum associated with the atomic transition in (d) occurs in emission as in absorption, and the process illustrated in (c) may yield strong continua extending over wide wavelength ranges. This type of emission continuum is used as a source of continuous radiation in the ultra-violet, where the intensity of the black-body spectrum from thermal sources becomes very low (Chapter 4). The lower state may be either weakly bound by van der Waals forces, as in (c), which is the case for the inert gas continua, or it may be a repulsive (anti-bonding) state, which is the case for the H_2 continuum.

3.8.3 PREDISSOCIATION

If the potential curve for a stable electronic state A is crossed by that for a state B, which is either unstable or has a lower dissociation energy (Fig. 3.13(a) and (b)), the system can, subject to certain selection rules, make a radiationless transition from A to B, followed by dissociation. At the particular internuclear distance R_P, corresponding to the crossing point P, this probability can be very high – up to 10^{11} s^{-1}.

The effect on the absorption spectrum from a third state X is that the A–X band system becomes diffuse for v' levels close to the intersection, where the upper level is described by a mixture of states A and B. The lifetime for dissociation from A via B is determined by the amount of admixture of B, and the broadening of the rotational lines becomes apparent when the 'uncertainty principle broadening' exceeds the Doppler width. The effect is entirely analogous to the autoionization broadening discussed in section 2.10. If the lifetime against radiationless transitions is $\sim 10^{-11}$ s, which is of the order of the time taken for a rotation, the rotational line structure is entirely smeared out for these particular v' levels. It may become sharp again for higher as well as lower values of v'. In emission the high probability of dissociation as against radiation ($\sim 10^8$ s^{-1}) results in the weakening or complete disappearance of bands with these critical values of v'.

Predissociation is just a particular form of configuration interaction, and the same selection rules apply (compare section 2.10). The angular momentum, (+, −) and (s, a) properties are as well defined for an unstable as for a stable molecular state, and the selection rules for radiationless transitions are identical

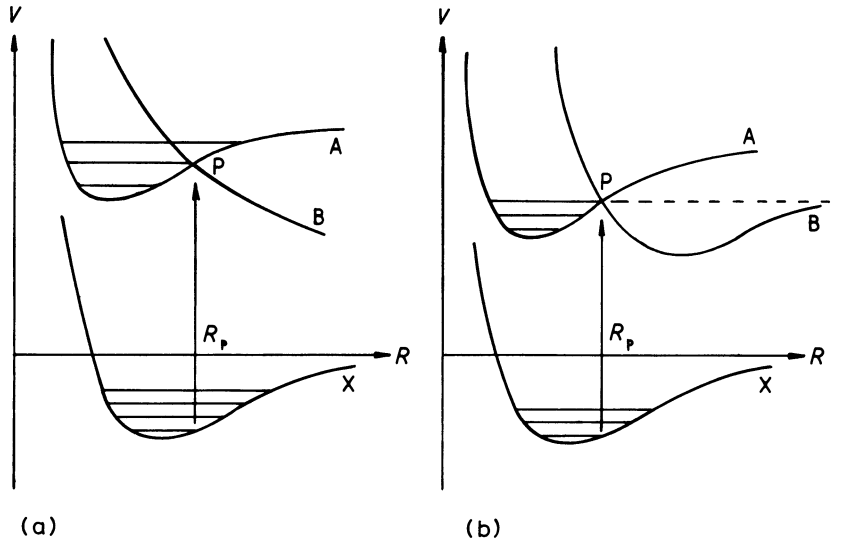

Figure 3.13 Pre-dissociation in the transition X→A. State A is crossed by a state B which either (a) is unstable or (b) has a lower dissociation energy. P is the energy at which A and B are degenerate.

to those for radiative transitions, with the sole exception that for homonuclear molecules both states must have the same nuclear interchange symmetry g or u, as against the electric dipole transition rule g ↔ u. This is entirely analogous to the situation for atoms, where configuration mixing occurs only between states with the same parity, whereas electric dipole radiation requires opposite parity.

3.9 Remarks on polyatomic molecules

The structure and spectra of polyatomic molecules are beyond the scope of this book. In this section we point out a few characteristics that can be understood from what has been said about diatomic molecules. For a full treatment a reference such as [2, 5, 9] should be consulted, and shorter accounts may be found in [1, 6].

The importance of symmetry considerations has already been made apparent in the discussion of diatomic molecules, and they are central to the treatment of polyatomic molecules. The symmetry of the wave functions determines both the interactions that contribute to the energy, via integrals of the form $\int \psi^* \mathbf{H} \psi \, d\boldsymbol{r}$, and the selection rules via integrals of the form $\int \psi_1^* e\boldsymbol{r} \psi_2 \, d\boldsymbol{r}$, since in each case the integral vanishes if the integrand is not symmetric over all space. Molecules are grouped according to their symmetry properties, and group theory is then

used to evaluate the result of operating with, for example, a particular transition operator.

Starting with rotation, one can define, as for any rigid body, three principal axes of rotation. The moments of inertia about these three axes are related if the molecule fulfils certain symmetry requirements. Symmetric tops have two of them equal, and spherical tops have all three equal. In fact the spherical symmetry of the latter means that it has no permanent electric dipole moment and hence no pure rotational spectrum. Unfortunately, most molecules are asymmetric tops, but many of them are sufficiently close to symmetric tops for this class to be a good approximation.

The number of degrees of freedom in an n-atom molecule is $3n$, one for each of the three translational co-ordinates of each of the atoms. In the molecule three degrees of freedom are used to specify the translation of the centre of mass and three for rotation, leaving $3n-6$ to specify the positions of the nuclei relative to one another. There are thus $3n-6$ vibrational modes to be considered. (With a linear molecule, as with a diatomic molecule, rotation about the internuclear axis does not change the nuclear co-ordinates, and so only two degrees of freedom are required for rotation, leaving $3n-5$ for vibration.) It is possible to choose a set of $3n-6$ (or $3n-5$) co-ordinates in such a way that each equation of motion involves one co-ordinate only; quantum-mechanically, Schrödinger's equation can be separated into $3n-6$ independent equations in the same way. Each equation describes a 'normal mode' of vibration, which to a first approximation is simple harmonic with a characteristic frequency ν_k. The energy associated with each of these modes is separately quantized: $E_k = h\nu_k(v_k + \frac{1}{2})$. The actual motion of each nucleus is a complicated superposition of all the normal modes, but in simple molecules it is often possible to regard each normal mode as representing either a change in length of the bond between one pair of atoms (stretching vibration) or a change in the angle between two bonds (bending vibration). In more complex molecules the normal modes may describe vibrations either of the whole 'molecular skeleton' or of particular groups of atoms such as OH, NH and CH_3. Vibrations of the second type are characteristic of the group and almost independent of the particular molecule to which it is attached. In general some of these modes involve a change in electric dipole moment of the molecule, and thus give rise to vibrational spectra (infra-red active) and some do not (Raman active: section 9.10). All modes appear as band structure on electronic transitions in the visible and ultra-violet.

The vibrational frequencies of polyatomic molecules are in the same energy range as those of diatomic molecules – a few hundred to a few thousand cm^{-1}. The vibrational bands have a fine structure of rotational transitions, as with diatomic molecules, and indeed the band structure of small (triatomic) molecules is rather similar to that of diatomic. However, since large molecules tend to have large moments of inertia and hence small rotational constants B, the separation of the rotational lines may become comparable with their intrinsic width,

resulting in a quasi-continuum. This effect may be compounded by the overlapping of different vibrational modes.

References

FURTHER GENERAL READING (in alphabetical order)

1. Banwell, C.N. (1966) *Fundamentals of Molecular Spectroscopy*, McGraw-Hill, New York.
2. Barrow, G.M. (1962) *Introduction to Molecular Spectra*, McGraw-Hill, New York.
*3. Bransden, B.H. and Joachain, C.J. (1983) *Physics of Atoms and Molecules*, Longmans, London.
4. Coulson, C.A. (1979) *Valence*, 3rd edn, Oxford University Press.
5. Hollas, J.M. (1987) *Modern Spectroscopy*, Wiley, Chichester.
6. Steinfeld, J.I. (1974) *Molecules and Radiation*, MIT Press, Cambridge, Massachusetts.

FOR A THOROUGH, COMPREHENSIVE TREATMENT

See the three volumes by G. Herzberg, published by Van Nostrand, New York.
7. *Spectra of Diatomic Molecules* (1959).
*8. *Infra-red and Raman Spectra* (1947).
9. *Electronic Spectra of Polyatomic Molecules* (1966).

The fourth volume in the series lists the spectroscopic constants for diatomic molecules:
10. Huber, K.P. and Herzberg, G. (1979) *Constants of Diatomic Molecules*, Van Nostrand, New York.

Detailed calculations of rotational structure are given in
11. Kovacs, I. (1969) *Rotational Structure in the Spectra of Diatomic Molecules*, Hilger, London.

Modern 'quantum chemistry' is reviewed by
12. McWeeny, R. and Pickup, B.T. (1980) Quantum theory of molecular electronic structure, *Rep. Prog. Phys.*, **43**, 1065.

For identification and photographs of actual spectra
13. Pearse, R.W.B. and Gaydon, A.G. (1976) *The Identification of Molecular Spectra*, 4th edn, Chapman and Hall, London.

* Further reading, not cited in text.

FOUR

Light sources and detectors

Light sources for spectroscopy may be divided into broad-band (or continuum) sources suitable for intensity standards or as backgrounds for absorption spectroscopy, and narrow-band (or line) sources for emission spectroscopy. Consideration of lasers is deferred to Chapter 8. Before discussing individual types of light source it is useful to define 'light intensity' and to deal briefly with some general concepts such as temperature, excitation energy and equilibrium.

4.1 Definitions of light flux

The quantitative description of optical radiation has been plagued by changes of terminology, by a wide variety of units, and by the ambiguity of words such as intensity which have both a general and a particular meaning. In this section we aim to assemble the minimum necessary set of definitions and to show the relations between them, using the currently accepted terminology [3].

If radiant power W (watts) falls on an area A (Fig. 4.1(a)), the *flux density* at the surface is given by

$$I = W/A \tag{4.1}$$

This quantity is now called *irradiance*. It is measured in units of $\mathrm{W\,m^{-2}}$ and may apply to a radiating surface, to the power crossing an imaginary surface in space, or to the power falling on a real surface. We denote it here by I rather than the 'preferred' E to avoid confusion with both energy and electric field.

I tells us nothing about the angular spread of the incident power. Consider now the power $\mathrm{d}W$ incident at an angle θ to the normal to the surface and contained within a solid angle $\mathrm{d}\Omega$ (Fig. 4.1(b)). The *radiance* L is defined as the power per unit projected area per unit solid angle:

$$L_\theta = \frac{\mathrm{d}W}{A\cos\theta\,\mathrm{d}\Omega} \tag{4.2}$$

This quantity was formerly called *brightness*, and its units are $\mathrm{W\,m^{-2}\,sr^{-1}}$. Again, it may apply to a source or to a real or imaginary detecting surface.

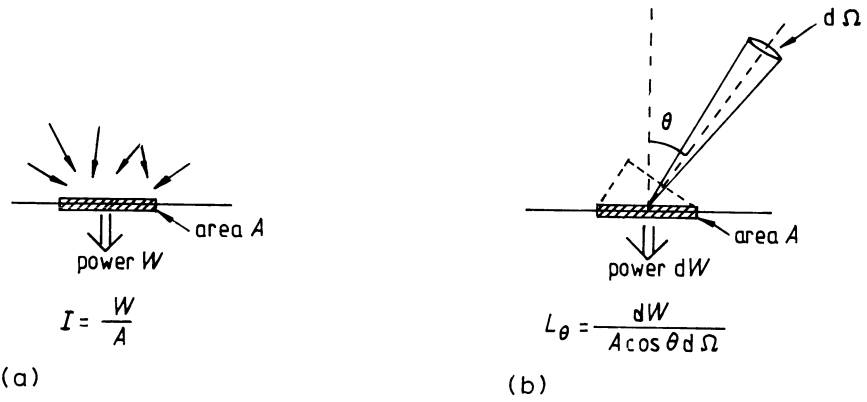

Figure 4.1 The definitions of (a) flux density (irradiance) I and (b) radiance L.

I is obtained from L by integrating over the relevant solid angle:

$$I = \frac{W}{A} = \int_\Omega \frac{dW}{A} = \int L_\theta \cos\theta \, d\Omega \tag{4.3}$$

If the radiation is *isotropic*, which means that L_θ has a constant value L independent of θ, then using $d\Omega = \sin\theta \, d\theta \, d\varphi$,

$$I = L \int_{\varphi=0}^{2\pi} d\varphi \int_{\theta=0}^{\pi/2} \cos\theta \sin\theta \, d\theta = 2\pi L \int_0^1 \sin\theta \, d(\sin\theta) = \pi L \tag{4.4}$$

(The spherical polar co-ordinate system is that of Fig. 2.1 with the normal as the z-axis.)

The familiar word *intensity* should properly be restricted to the power per unit solid angle radiated by a source, which makes it a property of the source rather than of the radiating surface. For a small radiating surface of area A we can write

$$\text{intensity} = LA \quad \text{W sr}^{-1}$$

but this is strictly true only for a point source, and intensity is not a useful quantity for discussing the interaction between radiation and matter.

The adoption of two such similar names as radiance and irradiance for two rather easily confused quantities appears to the occasional user to be a curious and unnecessary complication. In what follows we shall use radiance but avoid irradiance wherever possible. In fact, radiance is the more useful of these quantities in the context of optical systems because it has the same value throughout the system so long as power is not lost from the beam by absorption or scattering. Constant radiance is a consequence of constant throughput, defined as the product of the cross-sectional area A of the light beam and the solid angle delimiting it. Figure 4.2 shows this for a simple lens system, though it can be proved quite generally. The linear magnification here is v'/v, so

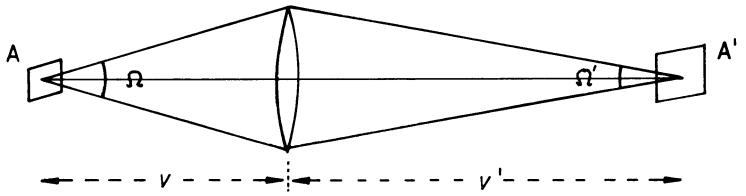

Figure 4.2 Conservation of throughput: $A\Omega = A'\Omega'$.

$A'/A = (v'/v)^2$, whereas $\Omega'/\Omega = (v/v')^2$. Therefore

$$\text{throughput} = A\Omega = A'\Omega' \tag{4.5}$$

From equation 4.2, with A and A' normal to the optic axis, we have $W = LA\Omega$ and hence, if W is conserved,

$$L' = L \tag{4.6}$$

It should be noted in passing that a further string of names and units attaches to these three basic quantities in measurements of visible light flux. For example, radiance becomes luminance and is measured in lumens $m^{-2}\,sr^{-1}$ or candela m^{-2}. There are 680 lm per watt at 555 nm, the peak sensitivity of the human eye. These units are not used in this book.

If one considers radiation from a thermodynamic rather than a radiometric starting point, the basic quantity is *energy density*, ρ, measured in W m^{-3}. The energy density associated with an electric field is given by $\rho = \frac{1}{2}\boldsymbol{D}\cdot\boldsymbol{E}$, where $\boldsymbol{E}$ and $\boldsymbol{D}$ are the electric field and displacement, respectively. The magnetic field of the electromagnetic wave contributes an equal amount of energy, so for an unpolarized electromagnetic wave in vacuum

$$\rho = \varepsilon_0 E_0^2 \tag{4.7}$$

where E_0 is the amplitude of the electric field vector and ε_0 is the permittivity of vacuum. If the light is plane polarized a factor of $\frac{1}{2}$ is required.

To relate ρ to the radiometric quantities defined above, consider first a collimated beam of cross-sectional area A. All the energy in a layer of thickness ct (where c is the speed of light) will cross A in time t – i.e. $Act\rho$ J. The power, or flux W, crossing A is thus $Ac\rho$ W, and the flux density (irradiance) is

$$\text{(collimated radiation)} \quad I = \rho c \tag{4.8}$$

I is related to the radiance in the direction of the beam by equation 4.3, giving

$$I = \int_{\Omega} L \, d\Omega = \rho c \tag{4.9}$$

The angular spread $d\Omega$ of a collimated beam is never zero: even in a 'perfect' collimator there has to be a diffraction spread, so although L may be large in the direction of the beam it is never infinite. At the other extreme, if the radiation is

isotropic, the radiance in all directions is by definition the same, and equation 4.9 becomes

$$L \int_{\text{sphere}} d\Omega = \rho c$$

therefore

$$\text{(isotropic radiation)} \quad L = \rho c/4\pi \tag{4.10}$$

Finally, the subject matter of this book is usually concerned with the spectral distribution of all these quantities; this is specified by inserting the word 'spectral' into the name – spectral radiance, spectral energy density, etc. Another potential confusion arises at this point, because a power distribution can be expressed either as power per unit wavelength interval or as power per unit frequency interval, and both of these are used in practice. From $v = c/\lambda$ we have

$$|dv| = \frac{c}{\lambda^2}|d\lambda|$$

Thus, one unit wavelength interval ($d\lambda = 1$) is c/λ^2 unit frequency intervals, and there is c/λ^2 times as much power per unit wavelength interval as there is per unit frequency interval.

4.2 Blackbody radiation and sources of continuum

For a solid or for a gas in thermodynamic equilibrium, the absolute temperature T can be defined thermodynamically. The spectral energy density of radiation in the gas or in a cavity in the solid is given by the Planck 'blackbody' distribution, which can be expressed either as energy per unit wavelength interval:

$$\rho(\lambda, T) = \frac{8\pi hc}{\lambda^5}\frac{1}{e^{hc/\lambda kT} - 1} \quad \text{J m}^{-3}\,\text{nm}^{-1} \tag{4.11a}$$

or as energy per unit frequency interval:

$$\rho(v, T) = \frac{8\pi hv^3}{c^3}\frac{1}{e^{hv/kT} - 1} \quad \text{J m}^{-3}\,\text{Hz}^{-1} \tag{4.11b}$$

where k is Boltzmann's constant.

The radiation in such a cavity is necessarily isotropic, so equation 4.10 can be used to give the radiance of a blackbody source, again on either a wavelength or a frequency scale:

$$L^{\text{B}}(\lambda, T) = \frac{2hc^2}{\lambda^5}\frac{1}{e^{hc/\lambda kT} - 1} \quad \text{W m}^{-2}\,\text{sr}^{-1}\,\text{nm}^{-1} \tag{4.12a}$$

$$L^{\text{B}}(v, T) = \frac{2hv^3}{c^2}\frac{1}{e^{hv/kT} - 1} \quad \text{W m}^{-2}\,\text{sr}^{-1}\,\text{Hz}^{-1} \tag{4.12b}$$

It has to be assumed that the radiation escapes from a hole small enough not to disturb the thermodynamic equilibrium.

Figure 4.3 shows the Planck distribution for two different temperatures. It does not matter whether we plot ρ or L as ordinate, one being simply proportional to the other, but it *does* matter whether the abscissa is λ or ν because, as seen at the end of the previous section, the ratio $\rho(\lambda)/\rho(\nu)$ is wavelength-dependent. At short wavelengths the $\rho(\nu)$ curve (Fig. 4.3(b)) falls off more rapidly than the $\rho(\lambda)$ curve (Fig. 4.3(a)). Similarly, for any given temperature the maxima of the two curves occur at different wavelengths. The broken line in Fig. 4.3(a) shows the locus of the maximum for different values of T. The shift to shorter wavelengths as T increases is expressed by Wien's displacement law:

$$\lambda_m T = \text{const.} = 2.884 \times 10^{-3}\ \text{m K} \tag{4.13}$$

λ_m here refers to the maximum of the $\rho(\lambda)$ distribution. One can bring out the difference in the two distributions rather clearly by noting that λ_m is given to a very good approximation by $hc/\lambda_m kT = 5$, whereas the maximum in the frequency distribution $\rho(\nu)$ is given by $h\nu_m/kT = 3$, which amounts to increasing λ_m by a factor of $\frac{5}{3}$ for a given T.

Two approximations to the Planck function are often useful. The Rayleigh–Jeans formula is valid when $h\nu \ll kT$:

$$L^{RJ}(\nu, T) = \frac{2kT\nu^2}{c^2} \tag{4.14}$$

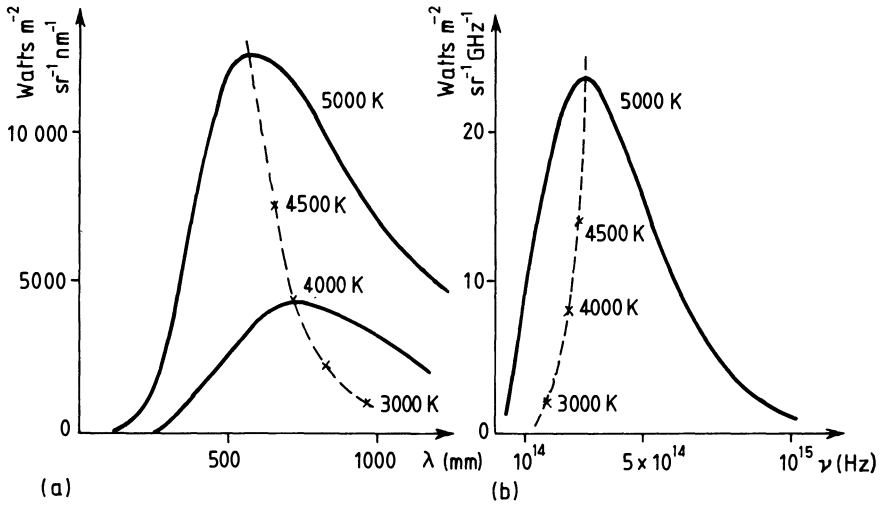

Figure 4.3 Spectral distribution of blackbody radiation (a) as a function of λ and (b) as a function of ν. The broken lines show the positions of the maxima at different temperatures. The two curves have different shapes and are not mirror images of one another.

and the Wien approximation when $h\nu \gg kT$:

$$L^{W}(\nu, T) = \frac{2h\nu^3}{c^2} e^{-h\nu/kT} \tag{4.15}$$

In the visible and ultra-violet regions Wien's approximation is valid for most laboratory sources. For $\lambda < 600$ nm the correction is within 5% for $T \leqslant 7600$ K, and the criterion for a 1% error is

$$\lambda\,(\text{nm}) \leqslant 3 \times 10^6 / T\,(\text{K})$$

A 5% accuracy for the Rayleigh–Jeans expression requires $\lambda > 14$ μm even for T as high as 10 000 K, so its applicability really is limited to the far infra-red.

No equilibrium source can have greater radiance at any wavelength than a black body at the same temperature, a fact that may be expressed by writing

$$L(\lambda, T) = \varepsilon_\lambda L^{B}(\lambda, T) \tag{4.16}$$

where the emissivity $\varepsilon_\lambda \leqslant 1$. ε is in general a function of wavelength, but if it is almost constant the source is known as a grey body.

The terms 'brightness temperature' and 'colour temperature' are sometimes used to specify the characteristics of near-blackbody radiation from a source of true temperature T_t. The brightness temperature T_b at any wavelength is the temperature of a black body radiating the same energy at that wavelength, i.e.

$$L(\lambda, T_t) = L^{B}(\lambda, T_b) = \varepsilon_\lambda L^{B}(\lambda, T_t)$$

$\varepsilon_\lambda < 1$ implies $T_b < T_t$, and T_b varies with λ unless ε is constant. The colour temperature T_c of a source is the temperature of a black body having the same *relative* radiance at two different wavelengths λ_1 and λ_2, so that

$$\frac{L_1(\lambda_1, T_t)}{L_2(\lambda_2, T_t)} = \frac{L^{B}(\lambda_1, T_c)}{L^{B}(\lambda_2, T_c)} = \frac{\varepsilon_1 L^{B}(\lambda_1, T_t)}{\varepsilon_2 L^{B}(\lambda_2, T_t)}$$

by using equation 4.16. For a grey body ($\varepsilon_1 = \varepsilon_2$) the colour temperature is identical to the true temperature. Lamp filaments are usually nearly grey, and this makes them satisfactory intensity standards for any application in which relative rather than absolute intensities are required.

The simplest background source for absorption spectroscopy is an incandescent solid, but solids necessarily have an upper temperature limit of about 3650 K, the melting point of tungsten. The peak wavelength for such a source is in the red, and its power falls rapidly in the ultra-violet. Sources of continuum for the far infra-red and the ultra-violet are described in section 4.9. For the most part they are not in thermodynamic equilibrium at a well-defined temperature, and their spectral energy distribution may bear little relation to the Planck function. The same is true of the pulsed sources used to observe absorption in very hot gases or plasmas, and of course lasers and masers are extreme examples of non-equilibrium sources.

4.3 Intensity standards

A primary standard is a source for which the radiated power is known theoretically as a function of wavelength. There are only two types of source for which this is true, a blackbody and an electron synchrotron. A true blackbody enclosure ('hohlraum') is closely approximated by a furnace with a small observation hole, and the ultimate standard of radiance is such a furnace at a temperature defined by the melting point of gold. It can be seen from Fig. 4.3 that a blackbody furnace is a feasible laboratory standard in the near infra-red, close to the peak of the distribution for temperatures between 1000 and 2000 K, but the sharp drop on the short wavelength side of the peak makes it highly inconvenient for the visible and ultra-violet regions. Blackbody radiation at higher temperatures (and thus at shorter wavelengths) has been realized in optically thick lines in high current, wall-stabilized arcs [20] (see chapter 11 for discussion of optical depth). The electron synchrotron is an even more inaccessible standard than a temperature-controlled furnace or a high-current arc. A brief account of this source is given in section 4.9.3, and further description can be found in [18, 19]. The peak wavelength, which is energy-dependent, is always in the far ultra-violet, but the power radiated at longer wavelengths is still sufficient for calibration against a 'transfer standard'.

Secondary standards that can be calibrated against the primary standard are essential. The most useful from the near infra-red to the near ultra-violet regions are tungsten and tungsten-halide filament lamps, which are fairly good grey bodies. When fitted with quartz envelopes they can be run up to a colour temperature of 3200–3300 K, giving a usable output from the infra-red to about 300 nm. For practical purposes such lamps, once calibrated, can be regarded as primary laboratory standards. For transfer standards between the tungsten lamp and the synchrotron or 'blackbody arc', two sources have been developed and evaluated, at both the National Bureau of Standards in Washington and the National Physical Laboratory in London. The first is the deuterium lamp [21]. The radiation consists mainly of a molecular recombination continuum analogous to that described in section 3.8 for hydrogen. The 'standard' lamp runs at 300 mA (30 W) and gives a fairly uniform continuum from 350 nm to about 165 nm, where it starts to be overlaid by molecular band structure. Less convenient but brighter by one to two orders of magnitude is the argon mini-arc [22]. This is a wall-stabilized arc of the type used in measurements of transition probabilities (section 12.3). Run at currents of up to 40 A at a pressure of 1–2 atm, it gives an almost line-free continuum between about 350 and 90 nm, arising from the recombination of argon ions and electrons; the axial temperature is 11 000–13 000 K. Its spectral radiance is said to be both reproducible and predictable from the discharge parameters.

As an alternative to an intensity standard, it is sometimes possible to calibrate some arbitrary but reproducible source by means of a detector which either measures absolute radiant power or has in its turn been calibrated against a

standard source. The calibration must include the spectrometer as well as the actual detector.

4.4 The excitation of line spectra

The power radiated by an atom or molecule at frequency ν_{12} as a result of a radiative transition between two discrete states (2) and (1) is determined by the probability of finding the atom or molecule in the initial state (2) as well as by the inherent probability of the particular transition $2 \rightarrow 1$. The latter is discussed in terms of time-dependent perturbation theory in section 2.8, and we now need to consider the former point – the distribution of atoms and molecules over the various excited states. 'Atoms' in what follows should be taken to apply equally to molecules, provided that the rotational levels are regarded as separate and discrete.

It can be shown by statistical mechanics that for a system in thermodynamic equilibrium at temperature T the ratio of the numbers of atoms occupying the two energy states E_2 and E_1 is given by Boltzmann's equation:

$$\frac{N_2}{N_1} = \frac{g_2}{g_1} \mathrm{e}^{-(E_2 - E_1)/kT} \tag{4.17}$$

The statistical weight g of a level is equal to its degeneracy. The degeneracy of a level with total angular momentum quantum number J is $2J+1$ in the absence of an external field (section 2.7), but in a magnetic field each J-level is split into $2J+1$ sub-levels, each with a different value of M and unit statistical weight. Similarly, each rotational level of the simple vibrator–rotator diatomic molecule has a statistical weight of $2J+1$. Vibrational levels are non-degenerate.

The conditions in which Boltzmann's relation may be expected to hold require some discussion. A source of line spectra cannot be in complete thermodynamic equilibrium, because if the radiating gas is enclosed in a blackbody cavity its spectrum is described by the Planck distribution and is necessarily continuous. What, then, is meant by the temperature of a gas in incomplete thermodynamic equilibrium? It is shown in Chapter 13 that the Boltzmann relation holds in a condition known as *local thermodynamic equilibrium*, commonly abbreviated to LTE, in which atomic states are populated and depopulated predominantly by collisions rather than by radiation. T is then the parameter determining the mean kinetic energy, $\frac{3}{2}kT$, of the colliding particles, which in LTE must necessarily follow a Maxwellian velocity distribution. In practice (see section 10.6 for justification) electrons are so much more effective for collisional excitation than are atoms or ions that only the electron velocity distribution is relevant, and T is therefore T_e, the 'electron temperature'. Indeed, it is often expressed in electron volts rather than degrees; since $1\ \mathrm{eV} = 1.6 \times 10^{-19}\ \mathrm{J}$, and $k = 1.38 \times 10^{-23}\ \mathrm{J\ K^{-1}}$, a useful equivalent for order-of-

magnitude calculations is that 1 eV corresponds to about 10 000 K (the exact value is 1.16×10^4 K).

The ratio g_2/g_1 in equation 4.17 is of order unity for atoms, so the likelihood of populating the excited state (2) depends on its excitation energy E_2 relative to kT. For spectroscopic purposes k may be conveniently expressed in wavenumbers as 0.7 cm^{-1} per degree. At room temperature kT is about 200 cm^{-1}, and as already noted one would expect the rotational and perhaps a few of the vibrational levels of the ground electronic state of a molecule to be populated. Most excited electronic states of molecules and atoms are some tens of thousands of wavenumbers above the ground state and require electron temperatures of several thousand degrees for appreciable population. For the yellow resonance lines of sodium, for example, $E_2 \simeq 17\,000$ cm^{-1} and $e^{-E_2/kT}$ is only about 0.1 even at 10 000 K.

The excitation of ionic spectra depends on the degree of ionization in the source. In LTE, when ionization and the inverse process of recombination are predominantly due to electron collisions, the ratio of ions to neutral atoms is given by Saha's equation, equation 13.21. This ratio has a certain similarity to the N_2/N_1 ratio of Boltzmann's equation, in that it is proportional to $e^{-\chi/kT}$ where χ is the ionization energy; but it is also inversely proportional to the electron density, and there are additional temperature-dependent terms. The effect of these is that there may be appreciable ionization even if χ is much larger than kT (section 13.5). In conditions typical of an arc or shock tube – 5000 K with an electron density of 10^{15} cm^{-3} (10^{21} m^{-3}) – the degree of ionization of calcium ($\chi = 6.1$ eV) is about 70%, while that of oxygen ($\chi = 13.6$ eV) is almost zero. At 10 000 K and the same electron density, ionization is virtually complete for calcium and is about 40% for oxygen. Multiple ionization may become significant at this sort of temperature, and about 60% of the calcium would be doubly ionized in this particular example.

4.5 The Einstein probability coefficients

It is shown in section 2.8 that the probability of a transition between two states due to an electromagnetic wave can be calculated from time-dependent perturbation theory if the wave functions of the two states are known. But this treatment says nothing about *spontaneous* emission, which we know to occur in the absence of resonant electromagnetic radiation and which we might expect to be related in some way to absorption. In fact, the relation between the probabilities for absorption and emission was first derived by Einstein from thermodynamical considerations before anything was known about atomic structure. We derive these relations here.

Figure 4.4 shows the energy levels E_1 and E_2 of an atom (or molecule) with populations N_1 and N_2, respectively. There are three possible radiative processes connecting the two levels, two obvious and one less so. First, an atom in

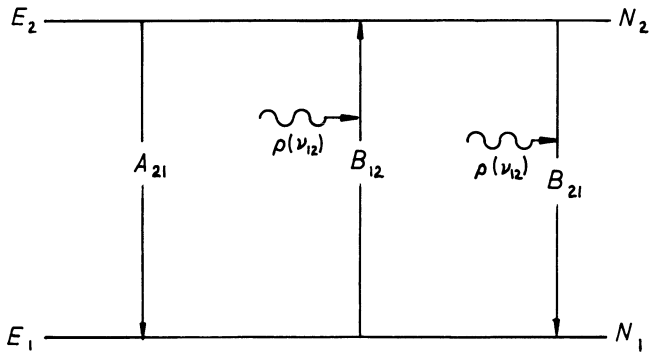

Figure 4.4 The Einstein probability coefficients. N_1 and N_2 are the population densities of the levels of energy E_1 and E_2, and $\nu_{12}=(E_2-E_1)/h$.

level 2 may undergo spontaneously a transition to level 1 with emission of energy $h\nu_{12}$; the probability of this occurring in unit time is defined by the coefficient A_{21}. The number of such transitions per second is therefore $A_{21}N_2$. Secondly, in the presence of radiation of density $\rho(\nu_{12})$ of the appropriate frequency an atom in level 1 may jump to level 2 with absorption of energy $h\nu_{12}$; the probability of this process is defined by $B_{21}\rho\,\mathrm{s}^{-1}$. Finally, an atom in state 2 in the presence of radiation $\rho(\nu_{12})$ may undergo *stimulated* emission to level 1 with emission of energy $h\nu_{12}$; the probability of this is defined by $B_{21}\rho$, and it gives rise to $B_{21}N_2\rho$ additional downward transitions per second. This last process is the less obvious one. Its existence is a necessary consequence of the Planck distribution, and quantum-mechanically it arises from exactly the same time-dependent perturbation as absorption. It is often helpful to regard it as negative absorption because the emitted radiation has exactly the same direction and phase as the stimulating radiation.

To obtain the relations between the coefficients, assume the assembly of atoms to be enclosed in a blackbody cavity in thermodynamic equilibrium at temperature T. The population ratio N_2/N_1 and the radiation density $\rho(\nu_{12})$ are given by the Boltzmann and Planck relations, respectively (equations 4.17 and 4.11). In equilibrium the rate at which atoms enter level 2 must be equal to the rate at which they leave it. By the principle of detailed balance (section 13.8) this equality must hold separately for collisional and radiative processes and for transitions to and from all other levels individually. In particular, the rate of radiative transition downward $2\rightarrow1$ must equal the rate upward $1\rightarrow2$:

$$A_{21}N_2+B_{21}\rho(\nu_{12})N_2=B_{12}\rho(\nu_{12})N_1 \tag{4.18}$$

Rearranging this

$$\rho(\nu_{12})=\frac{A_{21}N_2}{B_{12}N_1-B_{21}N_2}=\frac{A_{21}}{B_{12}(N_1/N_2)-B_{21}}$$

By Boltzmann's equation (equation 4.17)

$$\frac{N_1}{N_2}=\frac{g_1}{g_2}\,e^{h\nu_{12}/kT}$$

where E_2-E_1 has been set equal to $h\nu_{12}$. Therefore,

$$\rho(\nu_{12})=\frac{A_{21}}{B_{12}(g_1/g_2)e^{h\nu_{12}/kT}-B_{21}}$$

This is identical with the Planck expression

$$\rho(\nu_{12})=\frac{8\pi h\nu_{12}^3/c^3}{e^{h\nu_{12}/kT}-1}$$

for all values of T if and only if

$$\left.\begin{aligned} g_1B_{12}&=g_2B_{21}\\ A_{21}=\frac{8\pi h\nu_{12}^3}{c^3}B_{21}&=\frac{g_1}{g_2}\frac{8\pi h\nu_{12}^3}{c^3}B_{12}\end{aligned}\right\} \tag{4.19}$$

The Einstein coefficients are intrinsic atomic properties which can in principle be calculated from the wave functions, and they do not depend on the environment of the atoms. Hence, although the relations 4.19 were derived for thermodynamic equilibrium, they must still hold when the atoms are moved out of the blackbody enclosure – i.e. they are universally true.

The role of stimulated emission is made a little clearer by rewriting equation 4.18 as

$$\begin{aligned} N_2A_{21}&=N_1\rho\left(B_{12}-\frac{N_2}{N_1}B_{21}\right)\\ &=N_1\rho B_{12}\left(1-\frac{g_1}{g_2}\frac{N_2}{N_1}\right)\end{aligned} \tag{4.20}$$

which shows that it is effectively negative absorption, decreasing the actual absorption by the fraction g_1N_2/g_2N_1. In LTE this fraction is equal, by Boltzmann's relation, to $\exp(-h\nu_{12}/kT)$, which is negligibly small for $h\nu_{12}/kT\gg 1$. This inequality is just the condition for the validity of Wien's approximation (equation 4.15), so the stimulated emission can be held responsible for the -1 term in the Planck distribution; this is also clear from a comparison of the two expressions above for $\rho(\nu_{12})$.

Stimulated emission can be important even at short wavelengths if N_2 is greater than its LTE value. For $N_2g_1>N_1g_2$, stimulated emission exceeds absorption. This condition, known as population inversion, is necessary for laser action. Spontaneous emission is a loss mechanism for lasers, and since the ratio

of stimulated to spontaneous emission increases with λ^3, it is not surprising that masers were achieved before lasers.

The Einstein relations (equations 4.19) tell us nothing about the absolute values of the coefficients; but we already have a value for B_{12} (or B_{21}) for an electric dipole transition from equation 2.32:

$$P_{12}=\frac{1}{6\varepsilon_0\hbar^2}\rho(\nu_{12})|\langle 2|\boldsymbol{er}|1\rangle|^2$$

(The constant here differs by a factor 2π from that in equation 2.32 because unit angular frequency interval $=2\pi\times$ unit frequency interval.) This is a somewhat over-simplified expression in that no account has been taken of the degeneracy of either state; a more complete treatment is reserved for Chapter 11. However, the essential point is that the B-coefficients are proportional to $|\langle 2|\boldsymbol{er}|1\rangle|^2$, with analogous expressions for electric quadrupole, etc., transitions. They can therefore be calculated if the wave functions are known, and the A coefficient can then be found from equation 4.19. Spontaneous emission does not appear at all in the perturbation treatment. Direct calculation of A requires second quantization – that is, quantization of the radiation field as well as of the atomic energy levels – and is found to give exactly the same result.

There is an important pitfall to look out for in using the B-coefficients: they are sometimes defined in terms of *isotropic spectral radiance* rather than spectral radiation density. From equation 4.10, $L=\rho c/4\pi$ for isotropic radiation. Since the two definitions must give the same actual transition probability,

$$B^{\mathrm{L}}L(\nu_{12})=B^{\rho}\rho(\nu_{12})$$

leading to

$$B^{\mathrm{L}}=\frac{4\pi}{c}B^{\rho}\quad\text{and}\quad A_{21}=\frac{2h\nu_{12}^3}{c^2}B^{\mathrm{L}}_{12} \tag{4.21}$$

Actually the pitfall is even more extensive, because there is yet another variant: if ρ is defined in terms of *unit wavelength interval*, the relation $\rho(\lambda)=c/\lambda^2\rho(\nu)$ (section 4.1) gives

$$B^{\lambda}=\frac{\lambda^2}{c}B^{\rho}$$

The ambiguity of definition is confusing and regrettable. One must always make sure which convention is being used by any particular author.

4.6 The different types of light source

Apart from the temperature, or degree of excitation, of a light source, there are a number of other characteristics that may be important, such as the width of the

spectral lines emitted, the extent to which thermal equilibrium is attained, the spectral region in which the source is useful, whether it be pulsed or continuously running, and so on.

Line width forms the subject of Chapter 10. It will be seen that the two most important line-broadening processes are Doppler, due to the thermal motion of the emitting atoms, and pressure broadening. In low-pressure sources the first of these dominates, and at sufficiently high densities, particularly electron densities, pressure broadening becomes more important. In sources such as low-pressure glow discharges, which are not in local thermodynamic equilibrium, the gas kinetic temperature, which determines the Doppler width, may be close to room temperature while the electron temperature, which determines the excitation, may be several thousand degrees. In an arc at atmospheric pressure the two temperatures are usually similar (a few thousand degrees), and Doppler and pressure broadening may both be quite large.

Sources suitable for the measurement of transition probabilities, which have to satisfy particular requirements as to thermodynamic equilibrium, are further discussed in Chapter 12. The next few sections describe traditional spectroscopic emission sources and some recently developed devices, including sources suitable for the infra-red and the far ultra-violet. Since many of the new types of source do not appear in standard textbooks, a fairly extensive list of references is given at the end of the chapter. Discussion of masers and lasers is deferred to Chapter 8.

4.7 Traditional sources: flames, arcs, sparks and glow discharges

4.7.1 FLAMES

Most flames have temperatures of the order of 2000 K, although certain stoichiometric flames can reach about 4000 K. The low excitation energy ($\sim$0.25 eV) tends to produce a rather simple atomic spectrum, with the resonance lines dominating; indeed, the resonance lines of one of the alkali metals seeded into the flame is often used to measure its temperature by the reversal method (section 13.15.4). The principal uses of flames are for the study of the molecules and radicals formed during the combustion process, and for spectrochemical analysis by atomic absorption.

4.7.2 ARCS

The traditional type of free-burning arc runs typically at a few ampères and a few tens of volts, producing temperatures of the order of 5000 K; this is high enough to excite neutral spectra of atoms and molecules and a number of ionic lines. The arc is struck between metal or carbon electrodes, usually in air at atmospheric pressure, and has a column about 10 mm long. Where practical the arc pole pieces can be made of the metal whose spectrum is required, as in the

case of the iron and copper arcs used widely through the visible and quartz ultra-violet for reference wavelengths. Other elements may be introduced as salts in an arc run between carbon electrodes, or as trace impurities in copper electrodes. Arcs may also be run in gases other than air, and under reduced pressure.

Various special arcs have been developed to reach higher temperatures. High-current arcs (up to 500 A) with water-cooled electrodes can produce axial temperatures as high as 30 000 K. The arc column is restricted to a few mm diameter, either by a helical swirl of gas or water (vortex arc) or by a series of water-cooled diaphragms with axial holes (wall-stabilized arc). It is possible to introduce powdered solids as well as vapours into these arcs. Wall-stabilized arcs are particularly important in the measurement of transition probabilities, when high stability and a close approach to LTE are essential (section 12.3), and we have already referred to the use of the argon mini-arc as a laboratory intensity standard for the ultra-violet [22].

4.7.3 SPARKS

Sparks require high voltage (10–20 kV) from a transformer or induction coil to break down an insulating gap. The transformer may be connected directly across the spark gap, but normally it charges a capacitor, and the stored energy of a few joules is discharged across the gap every half-cycle. Several degrees of ionization can be reached in atoms of the metal forming the spark electrodes or of the gas through which it passes. The old names 'arc and spark spectra' for the spectra of neutral atoms and ions, respectively, come from these traditional methods of exciting them, although high-current arcs can in fact also produce multiple ionization.

In the vacuum ultra-violet ionization can be attained in a spark through a gas at low pressure confined in a capillary tube. The degree of ionization can conveniently be controlled by varying the gas pressure and the circuit inductance. Below 50 nm, and for very high ionization, a vacuum spark is necessary. The highest possible excitation is attained with an open spark at a voltage approaching 100 kV, when the electron temperature may be as high as 4 keV. In spite of the high electron density ($\sim 10^{20}$ cm^{-3}) the spark is too fast for full ionization equilibrium to be reached, and although 20 stages of ionization have been achieved, about 10 is more common. Rather less violent excitation is produced by a sliding spark, in which the voltage breakdown takes place across the surface of an insulator. The spectrum contains lines of electrode and insulator materials and of any low-pressure gas introduced into the gap.

4.7.4 GLOW DISCHARGES

Gases and vapours are often conveniently excited in the positive column of a glow discharge, or in an electrodeless discharge, at pressures of a few torr.

Molecular, atomic and ionic spectra may all be excited in this way. Excitation is by electron bombardment throughout the length of the positive column, which can be made to extend for a metre or so if necessary. A discharge of this type has to be run off a high voltage source (1 kV upwards) with a ballast resistance to stabilize it and to limit the current to something of the order of an ampère.

Another form of glow discharge, the hollow cathode, uses the cathode glow instead of the positive column. At a suitable pressure the cathode glow jumps into and completely fills a hole a few mm in diameter in the cathode, producing an intense and fairly compact light source. Material from the walls of the hole is sputtered into the discharge and there excited, so that metals and salts as well as gases can be studied. A choice of carrier gas from among the inert gases, or mixtures of them, allows the sputtering and the excitation to be varied, so that the spectra of ions or of neutral atoms can be observed. At one extreme the hollow cathode can be run up to an ampère or so to bring up weak spectral lines; at the other it can be run at a few mA with liquid nitrogen cooling to reduce the Doppler widths of the lines, and for high resolution work it has the further advantage that the electric field in the cathode is small so that Stark broadening and shift are also small.

Electrodeless discharges, run from radiofrequency or microwave sources, are particularly useful for producing high purity spectra of gases and vapours since there is no contamination from the electrode material.

4.8 Special sources

(See references at end of chapter for further details.)

4.8.1 SHOCK TUBES

Shock tubes are aerodynamic devices that have become useful spectroscopic sources. There are several different ways of producing a shock wave in a gas; for example, a pressure-driven shock is generated by bursting a diaphragm between a region of high-pressure light gas and one of low-pressure heavier gas. As the wave travels down the tube through the low-pressure gas it leaves behind it an almost uniform region of hot gas at a few thousand degrees and at a pressure of a few atmospheres, lasting, at any given observation point, for a few ms. The temperature and pressure range is comparable with that of an arc, but the shock tube has the advantage of a relatively large homogeneous region in which both temperature and pressure can be measured and independently controlled. Atomic species that do not form suitable volatile compounds can be introduced as powdered solids. The transit time is sufficiently long for easy time resolution, and the spectra of the constituent atoms or molecules can be studied either in emission or, using a flash as background, in absorption. The heating effect may be increased by reflecting the shock wave from the end of the tube.

4.8.2 PINCH DISCHARGES

The pinch discharge is another instance of a device developed for an entirely different purpose – in this case controlled thermonuclear reactions – that has found useful applications in spectroscopy. It works on the principle that the magnetic field associated with a pulse of high current can be used to squeeze a column of ionized gas rapidly in towards the axis. As the collapse is completed the ordered radial velocities are converted by collisions to random or thermal velocities, giving a narrow column of very hot gas lasting, typically, for a few microseconds before it expands and cools. The high degree of ionization attained in the pinch and its short duration make it comparable with a spark source, but with the advantages of much larger dimensions and greater reproducibility. The recombining plasma that follows the pinch also constitutes a useful source at a temperature of order 10 000 K and duration about 1 ms. Again, the fact that temperatures and electron densities can be measured independently is a great asset for studies of line broadening, forbidden lines, etc.

4.8.3 PLASMA FOCUS

This form of pinch discharge vies with the vacuum spark for the record for optical spectra of the shortest wavelength and the highest stage of ionization. The stored energy from a large bank of capacitors, amounting to several thousand joules, is discharged in a few microseconds, giving a peak current of order 10^6 A, the electrodes being coaxial cylinders arranged to form an open-ended 'plasma gun'. The discharge current is driven towards the open end, where it collapses to form a filament of extremely high energy density, with a temperature of order 1 keV and a volume of about 10 mm^3, lasting for only 100 ns. The electron density of about 10^{19} cm^{-3} is high enough to produce a high degree of ionization in this time. The spectra come from both the electrode material and the filling gas. For example, lines of the helium- and hydrogen-like spectra of 16 and 17 times ionized argon, at about 0.3 nm, have been photographed with this device.

4.8.4 LASER-PRODUCED PLASMAS

A plasma in the same temperature range as a pinch (50–500 eV), but with a very much higher electron density, $\sim 10^{20}$ cm^{-3}, may be produced by focusing a Q-switched ruby or neodymium laser on a solid target. Depending on the laser power, stages of ionization up to about 15 can be attained for a few ns. This type of excitation is obviously particularly appropriate for refractory solids.

4.8.5 INDUCTIVELY COUPLED PLASMAS

The ICP has effectively replaced the arc for spectrochemical analysis. A stream of argon at approximately atmospheric pressure forms a 'torch' excited by a

radiofrequency loop, into which can be injected either gases or solids in solution in the form of fine droplets produced by a nebulizer. The source is close to LTE at about 6000 K, appropriate for the spectra of neutral atoms and the first stage of ionization. Its merit for spectrochemistry is that it can be made accurately quantitative: the emission of a given species of atom depends linearly on its concentration in solution over a wide range.

4.8.6 BEAM FOIL SOURCE

Beam foil spectroscopy uses a source of an entirely different nature. A beam of ions accelerated to around 1 MeV is fired through a thin foil and emerges, by some process which is not entirely understood, as a mixture of ions of various degrees of ionization and in various states of excitation, but with its beam character unimpaired and its velocity essentially unaltered. A rich ionic spectrum is emitted in the region downstream of the foil as the ions return to their ground states. The degree of ionization reached may be up to about 7, and because of the non-thermal character of the excitation the spectra may differ considerably from those obtained in other sources. The intensity depends rather critically on the foil as to material, thickness and ageing; a light element is essential to reduce scattering and energy loss, and the commonest choice is carbon at a thickness of about 10 $\mu g\,cm^{-2}$ (say about 50 nm). There is sometimes difficulty in identifying the spectral lines, which are apt to suffer from severe blending because of the number of lines and their large Doppler widths (section 12.5.2), and as an aid to identification ions of different charge may be separated spatially with electric fields. The beam foil source has been found particularly useful in three areas: the analysis of ionic spectra, especially in following an iso-electronic sequence at intermediate degrees of ionization; the measurement of radiative lifetimes (Chapter 12); and experiments such as those involving optical pumping which exploit the low number density ($\sim 10^5\,cm^{-3}$) to eliminate collisions.

4.9 Absorption spectroscopy

4.9.1 GENERAL

Absorption spectroscopy requires a background source of radiation that is continuous over the relevant spectral region and an absorption cell. Except for the permanent gases, the latter is normally a furnace, running usually in an inert or reducing atmosphere. The temperature attainable is limited by the melting or softening point of the furnace material, which means in practice some 3000 K. Various flash-heating techniques have been developed for studying the absorption of short-lived molecular radicals (flash photolysis) and of very refractory solids (flash pyrolysis), and it is possible to use many of the emission sources described in the previous section in absorption, provided there is available a

background source of effectively greater brightness temperature – in practice some type of flash tube.

The 'cell' windows tend to present a problem in absorption spectroscopy: they are apt to get fogged with material sputtered or condensed on them, or to be chemically attacked by certain vapours, and in any case 110 nm is the lower-wavelength limit of transparency. Depending on the experiment, a few torr of buffer gas or a differential pumping system may be used as protection. For some metals over an appropriate temperature range a heat pipe is a very effective absorption cell: the metal vapour condensing at the cooler ends of the tube is returned to the hot centre by the capillary action of a suitable mesh, forming a sort of diffusion pump which pumps the buffer gas to the ends of the hot region. The metal vapour pressure is then determined entirely by the pressure of the buffer gas; winding up the power input to the furnace simply increases the length of the vapour column.

The simplest form of background continuum in the visible and near infra-red is the tungsten filament lamp, the characteristics of which are described in section 4.3. Higher temperatures, and therefore more blue and ultra-violet radiation, are produced by high-pressure gas arcs. The commercial version of the xenon arc, running at 50–80 atm pressure, has a colour temperature of about 6000 K and gives a usable continuum between about 750 and 190 nm. If higher brightness temperature is required, say to obtain an absorption spectrum from a hot plasma, a pulsed source is probably necessary. Pulsed xenon lamps are available commercially, as also are rather larger low-pressure flash tubes of the type described in section 4.9.3, reaching brightness temperatures up to 50 000 K in a flash of a few μs. This type of source is a suitable background for photolysis or shock-tube absorption spectroscopy.

4.9.2 THE INFRA-RED REGION

The intensity of the tungsten halogen filament lamp peaks at about 900 nm, and the quartz envelope ceases to transmit at about 3 μm. At longer wavelengths it is preferable to use a filament made of a mixture of rare-earth oxides operating at about 2000 K, known as the Nernst glower. At wavelengths above 10 μm this is beaten by the globar, a silicon carbide strip source, and by its more modern equivalent, a nichrome strip in a silica envelope; these are good greybodies from 2 to 40 μm. The most widely used source in the far infra-red is the mercury arc lamp at a pressure of a couple of atmospheres. Below about 60 μm, where the quartz envelope is opaque, this radiates as a blackbody, but at longer wavelengths the radiation is that of a plasma continuum in LTE at a few thousand degrees.

At still longer wavelengths, approaching 1 mm, we reach the microwave region, where the whole technique changes. The continuous source–disperser–non-selective detector combination is replaced by a narrow-band

radiator and tuned detector which are scanned together over the required spectral region.

4.9.3 THE ULTRA-VIOLET REGION

Continuous emission over the region from the quartz ultra-violet down to the resonance line of helium can be obtained in one way or another from discharges in hydrogen or the inert gases, although below 110 nm these must be run in windowless tubes. The high-pressure xenon arc referred to in section 4.9.1 is overtaken in radiance at the end of the quartz ultra-violet ($\sim$200nm) by the argon wall-stabilized arc and by the deuterium lamp (section 4.3). The laboratory counterpart of the latter, using hydrogen in a water-cooled discharge tube running at an ampère or so, radiates more power than the commercial deuterium lamp, but has the same short-wavelength limit of 160 nm.

Good continua extending some tens of nm to the long-wavelength side of their respective resonance lines can be obtained from all of the inert gases by running either a condensed or a microwave discharge at a pressure of a few hundred torr. The origin of these continua is the bound–free emission described in section 3.8.2. Between them, Xe, Ar, Ne and He cover the region from 180 to 60 nm.

The whole ultra-violet region is covered by the flash-tube source that was mentioned above as a general purpose source of high brightness temperature for the visible. In the original version devised by Lyman the flash was produced by discharging a capacitor through a capillary tube containing low-pressure gas. This had the disadvantages that the capillary was eroded rather rapidly and the continuous spectrum was overlaid by emission and absorption lines from the material scoured off the walls; the Garton modification reduces these problems by using a wider bore tube in a low-inductance circuit so that the necessary current density is achieved in a very short flash (~ 1 μs). A discharge of a few μF at 10 kV produces a continuum usable down to about 30 nm, but increasingly overlaid with lines towards the short-wavelength end. Shortwards of about 50 nm the best flash continuum is provided by the BVR source, called after its developers Balloffet, Vodar and Romand. This is a condensed spark in vacuum with a uranium pin acting as anode. The continuous radiation is emitted from a very small blob of hot gas just off the tip of the anode. Under typical operating conditions the flash from a 0.1 μF capacitor charged to 30 kV lasts about 1 μs and produces a continuum down to 1 nm or less.

Laser-produced plasmas have already been mentioned as emission sources. With suitable target material they can also act as reasonably line-free continua. The rare earths, probably because of the sheer complexity of their partly filled f- and d-shell configurations, have quasi-continuous ionic spectra all through the region 200 to about 4 nm. Typically, a joule of radiation from a ruby or neodymium laser is focused on the target to a spot less than 1 mm in diameter.

The ultra-violet radiation is emitted as a pulse lasting about as long as the laser pulse (5–30 ns) and is fairly reproducible from shot to shot.

A source of an entirely different type is the Bremsstrahlung radiation emitted by the electrons in a synchrotron accelerator or storage ring. The electrons are accelerated towards the centre of the circular orbit, and at low energies in accordance with classical electromagnetic theory they emit radiation in all directions perpendicular to the orbital radius. At high energies, when relativistic effects become important, the radiation is concentrated into a narrow cone in the direction of the instantaneous motion of the electron. The greater the energy, the smaller is the angle of this cone. The radiation is therefore virtually confined to the orbital plane, and it is plane polarized with the electric vector in this plane. The power radiated at any wavelength can be calculated from the radius of the orbit and the energy of the electrons, so that this source is in the same category as the blackbody as a true absolute intensity standard, albeit an even less accessible one to most experimenters. Figure 4.5 shows the general form of the spectral power distribution, which has something of the appearance of a set of blackbody curves, but with a different functional dependence on energy and wavelength. The important parameter is the 'critical wavelength', λ_c, which is given by $\lambda_c = 0.56\, R/E^3$ nm, where R is the radius of the orbit in metres and E is the energy in GeV. The peak of the distribution occurs at $0.4\lambda_c$ and therefore varies as E^{-3} for a given radius. The peak power varies as E^7 and the total

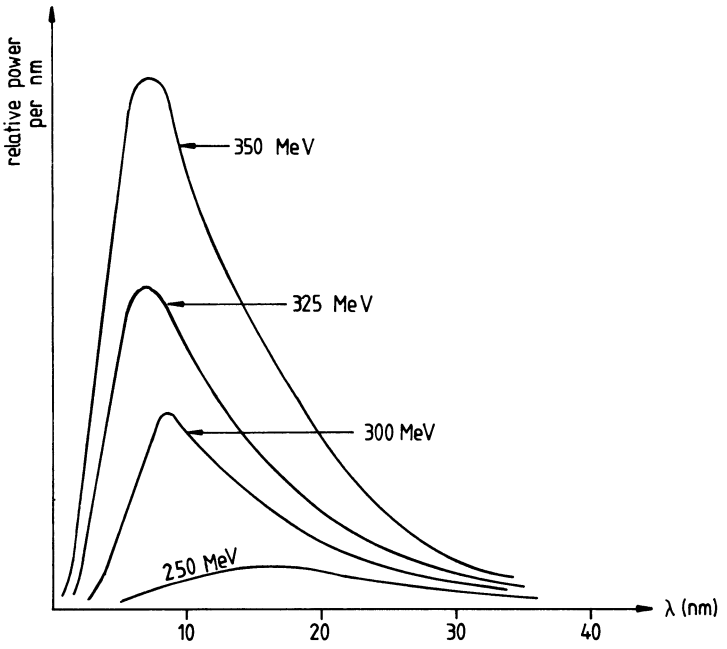

Figure 4.5 Spectral distribution of synchrotron radiation for different electron energies.

power as E^4, again for a fixed R. For $\lambda \gg \lambda_c$ the power is almost independent of E. Over the past 10 or 15 years synchrotron radiation has acquired so many important applications, in chemistry and biology as well as in several branches of physics besides spectroscopy, that a number of so-called storage rings have been built specifically to use the radiation rather than the fast electrons themselves; the injected electrons circulate in an ultra-high vacuum for a couple of hours, during which the power drops slowly as electrons are lost. This is an intrinsically 'quieter' source than a synchrotron, in which there is inevitably some pulse-to-pulse variation.

4.10 General remarks on detectors

Detectors may respond either to incident power or to incident photon flux. The first type – thermopiles and bolometers, for example – are non-selective in wavelength response, and the second type – photographic emulsions and photodiodes, for example – are selective. Although different photon detectors can be used from the far ultra-violet to about 15 μm, a natural division between types occurs at about 1 μm, corresponding to a photon energy of 1.2 eV. Below this wavelength photons have sufficient energy to eject an electron from a photocathode or to sensitize the silver halide grains of a photographic emulsion. Above it, they can modify the characteristics of a semiconductor; photon detection by photoconductors or semiconducting photodiodes is possible up to about 15 μm. At longer wavelengths certain photoconductors are still usable, but thermal detectors are more common. These, being relatively slow and insensitive, are not normally used at the shorter wavelengths except for absolute measurements.

The useful information obtained from any detector depends on the signal-to-noise ratio in the final output. Noise may originate from fluctuations of the light source itself or of the background radiation; from photon noise, associated with the random arrival of photons from a steady source; from the detector itself; or from the amplifying system that follows it. For the present purpose we shall ignore source fluctuations and look at the rest of the list. Background radiation makes an important contribution to noise in the infra-red because the spectrometer and its surroundings (including the spectroscopist) all radiate most effectively around 10 μm; the background can usually be ignored for $\lambda < 1$ μm. Detector noise comes from fluctuations in the dark current, which is the signal in the absence of incident radiation, and it is essentially thermal in origin. For the photoemissive detectors, working at cut-off energies much greater than kT (remember that $h\nu$ at 1 μm is 1.2 eV whereas kT at room temperature is 0.025 eV) the detector noise also is usually negligible. Amplifier noise may limit the performance of photodiodes at low light levels, but this limitation is avoided by photomultipliers, in which the amplification takes place in the vacuum tube itself with little or no degradation of the initial signal-to-noise ratio. In practice,

therefore, performance in the visible and ultra-violet is usually limited by the photon noise in the signal.

In the infra-red region the detector, amplifier and background noise may all be important. Dark current in the semiconducting devices is reduced by cooling, as also is the detector noise in thermal detectors, which arises from fluctuations in the 'dark' temperature of the detector. To reduce background noise and make a.c. amplification possible it is general practice to chop the light signal at some frequency f and use a phase-sensitive amplifier of bandwidth Δf. f is limited by the requirement that $1/f$ must be large compared with the response time of the detector, and it may range from 10 Hz for slow thermal detectors to a few kHz for photoconductors. The noise power associated with random fluctuations is approximately independent of frequency ('white noise'), so Δf sets the noise power passed by the amplifier; the corresponding noise in the output signal is proportional to the square root of this (see below).

It is useful to define here a few terms commonly used in the literature on detectors. The *responsivity* of a detector is defined by

$$R = V_0/W \quad \text{V W}^{-1} \tag{4.22}$$

where V_0 is the rms (root mean square) output voltage and W is the rms power input. The *noise equivalent power* (*NEP*) is defined as the input power that would be required to give an output voltage just equal to the rms noise voltage – in other words, the input power P_n required for a signal-to-noise ratio of unity. From equation 4.22, $P_n = V_n/R$, where V_n is the rms noise in the output voltage. The detectivity D is simply the inverse of the NEP, but the quantity much more widely used than D is the *specific detectivity* D^*, which is D normalized to a detector area of 1 cm^2 and a bandwidth of 1 Hz according to the relation

$$D^* = \frac{\sqrt{(A\Delta f)}}{P_n} \quad \text{cm Hz}^{1/2}\,\text{W}^{-1} \tag{4.23}$$

The logic of this apparently strange definition is that for many sources of noise the NEP is proportional to the square root of the detector area A, and for white noise it is also proportional to the square root of the bandwidth Δf. D^* therefore serves as a figure of merit for most types of detector at any particular wavelength and temperature. The square root dependence can be justified as follows. Dark current, or current due to background radiation, are both proportional to A. The *fluctuations* of these currents, according to Poisson statistics, are proportional to $\sqrt{A}$. Similarly, the rms value of the fluctuations, V_n, is proportional to $1/\sqrt{t}$, where t is the integration time; and since t is limited by the bandpass according to $t \sim 1/\Delta f$ it follows that V_n is proportional to $\sqrt{\Delta f}$, and

$$P_n \propto V_n \propto \sqrt{(A\Delta f)}$$

Having dealt at some length with noise, we now consider the signal levels recorded by a detector in association with a spectrometer. The dispersion of a grating is almost linear with wavelength, so a thermal detector with a grating

records W nm^{-1}. However, the dispersion of a Fourier transform spectrometer is effectively linear with frequency or wavenumber, and a thermal detector therefore records W cm^{-1}. The distinction is important because, as noted in section 4.2, the spectral radiance of a thermal source decreases towards short wavelengths faster (by a factor of $1/\lambda^2$) for the frequency scale than for the wavelength scale. Furthermore, the number of photons per watt is proportional to λ, so for a photon detector there is a further factor of $1/\lambda$ against short wavelengths. The position of the effective maximum of the blackbody distribution is a good illustration of this effect. Equation 4.13 gave Wien's displacement law for the W nm^{-1} distribution as $\lambda_m T = 2880$, with λ in μm and T in K. For the distribution in W cm^{-1} the constant is 5100, and for photons cm^{-1} it is 9030. The peak wavelength for a 5000 K blackbody (the Sun, say, or a 10 A arc), shifts from 580 nm to 1 μm to 1.8 μm for these three types of measurement.

The remainder of this chapter gives some more detail on detectors for the infra-red, meaning $\lambda > 1$ μm, and detectors for the visible and ultra-violet, $\lambda < 1$ μm. There is also a section on photographic emulsions, the characteristics of which do not fit happily into a discussion of photoelectric detectors, and finally a few remarks on array detectors which combine some of the advantages of both types of detector.

4.11 Detectors for the infra-red

These fall into two classes: photon detectors, which have a long-wavelength cut-off, and thermal detectors, which respond to radiative power irrespective of wavelength.

The photon detectors above 1 μm are all semiconductors of one kind or another. They may be formally divided into photoconductors, in which the conductance of the material is increased by the creation of a free electron (or hole) when a photon is absorbed, and p–n junction photodiodes, in which the current–voltage characteristics are changed by the hole–electron pair produced by absorption of a photon near the boundary layer. In fact, the second class may be used either in the photovoltaic mode as voltage generators or in the photoconductive mode as current generators, and we shall not worry about the distinction here. A wide variety of both types is available commercially with different spectral ranges, specific detectivities and time constants. A specialized text (e.g. [6, 7, 9]) or a commercial catalogue should be consulted for details.

Figure 4.6 gives a few typical curves of D^* as a function of wavelength, together with the curve for an ideal photoconductor and the straight line for an ideal thermal detector. 'Ideal' in this sense means background-limited at 300 K, and therefore depends on the acceptance angle of the detector. D^* has the same general spectral dependence for all real semiconductor detectors, rising to a peak approaching the ideal value and then dropping sharply. Most photodiodes (silicon, lead sulphide, lead selenium and indium antimonide) have cut-off

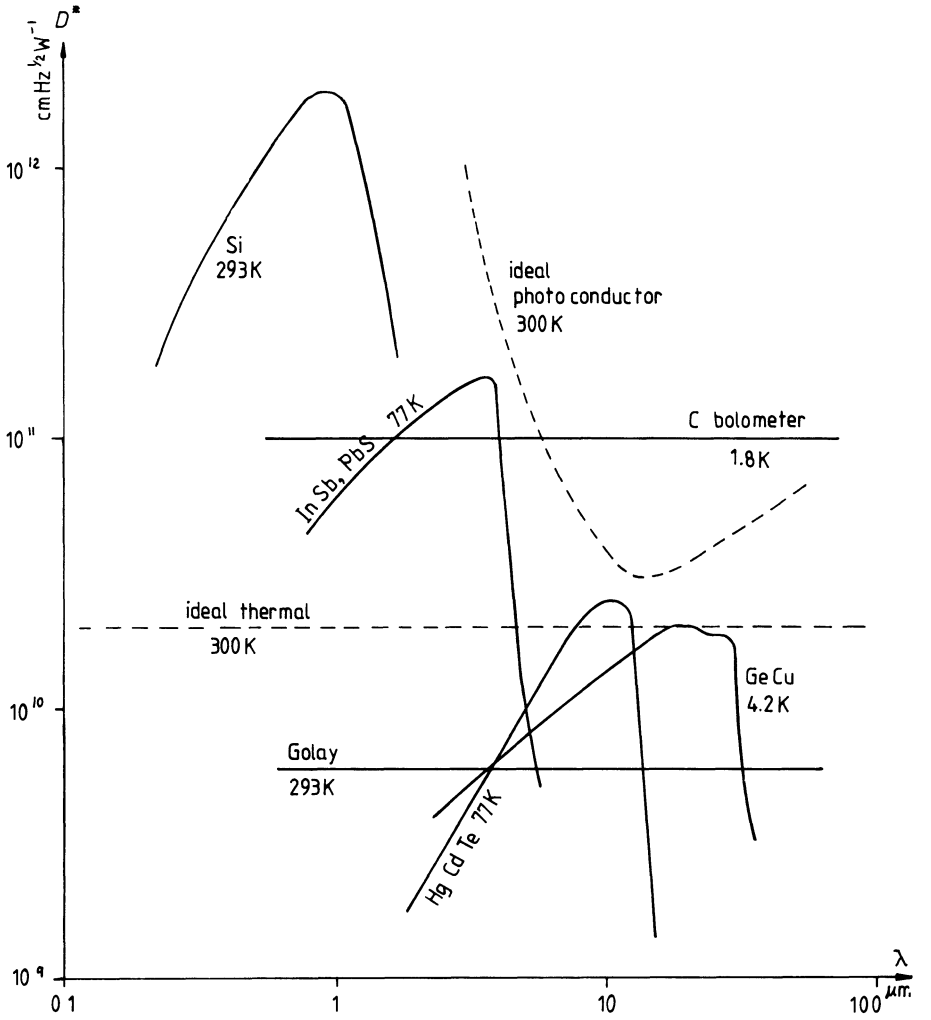

Figure 4.6 Specific detectivity D^* as a function of wavelength for some typical detectors. The curves for the ideal photoconductor and the ideal thermal detector refer to background-limited detectivity at 300 K.

wavelengths in the range up to 5 μm, but the various mercury–cadmium–telluride alloys work up to 15 μm. Peak values of D^* are mostly in the range 10^{10}–10^{11} cm $Hz^{1/2}$ W^{-1}, but absolute values must be treated with caution since they depend on temperature, acceptance angle, composition and (to a small extent) chopping frequency as well as cut-off wavelength. Extrinsic photoconductors, consisting of germanium doped with small amounts of impurity,

have cut-off wavelengths as high as 130 μm, but will work only with liquid-helium cooling. Indeed, In–Sb detectors can be used right up to the millimetre-wave region because the mobility of the free electrons is increased by the absorption of photon energy – i.e. the effective temperature of the electrons is raised above that of the crystal. For this reason they are known as 'hot electron' detectors.

The general form of the D^* curves can be understood if it is remembered first that any detector integrates background noise at all wavelengths up to its cut-off and, secondly, that the number of photons per watt increases linearly with λ. Thus, D^* for the ideal photoconductor drops with the increase in integrated background until the cut-off reaches about 10 μm, the 300 K peak, after which the further slow increase in integrated background is overtaken by the second effect. The approximately linear rise in D^* of the real detectors is also a consequence of the second effect: as λ increases, fewer watts are required to provide enough photons to equal the dark current. The response times of all these detectors depend on temperature and method of use as well as on type and area. They range from a few ns to about 100 μs.

Thermal detectors are most widely used in the far infra-red. The flat detectivity of the ideal thermal detector shown in Fig. 4.6 simply reflects its power – as opposed to photon – response, which integrates the entire background spectrum regardless of measurement wavelength. There are several different ways of measuring temperature changes. Thermopiles measure thermoelectric emf, bolometers measure resistance, pyroelectric detectors measure the change in polarization of a ferroelectric crystal, and the Golay cell measures the change in pressure of a gas cell. Of these, thermopiles are the most useful for absolute measurements at shorter wavelengths, but it is actually rather difficult to make them totally 'black' in the far infra-red. The other three types may be described briefly as follows.

Bolometers, measuring change in resistance from absorbed power, need to be cooled with liquid helium for high sensitivity. Either carbon or germanium may be used for the resistor, the latter having higher detectivity but a slower response time of several ms. Bolometers can be used all the way from the near infra-red to the millimetre-wave region and can be made to approach the 'ideal' background limit for D^*.

Pyroelectric detectors measure changes with temperature of the bulk dipole moment of a ferroelectric crystal, essentially by using the crystal as a capacitance in a suitable circuit. The emphasis is on the word 'changes': the incident radiation *must* be chopped for this type of detector, and the response depends on the chopping frequency. The detectivity is smaller by an order of magnitude than that of the carbon bolometer and the time constant is somewhat greater, but these detectors have the great advantage for many purposes of working at room temperature.

Finally, the Golay cell, another room-temperature device, is probably the most widely used far infra-red detector. The cell, containing xenon, has a

window at one end, a small flexibly mounted mirror at the other, and a thin film to absorb radiation in the middle. The heated gas deflects the mirror slightly, and the movement can be converted to an electrical signal with a suitable optical system and photocell. The time constants are such that the chopping frequency has to be about 10 Hz. Again, detectivity is a good order of magnitude below that of a cooled bolometer, but convenience rating is high.

4.12 Detectors for the visible and ultra-violet

In the region below 1 μm photoemission from a solid surface becomes possible. Vacuum photodiodes use this type of detection in its simplest form, measuring the current due to the electrons ejected from a suitable photocathode exposed to radiation. Current is proportional to photon flux over 6–8 orders of magnitude, and the useful spectral range is about 600–200 nm. In the visible and the near ultra-violet, down to about 300 nm, silicon p–n junction photodiodes as described in the previous section are usually preferable to vacuum photodiodes because of their smaller size and greater quantum efficiency ($\simeq 60\%$). Time constants can be in the ns range, and a typical detectivity curve is shown in Fig. 4.6.

At low light levels photodiodes tend to be limited by amplifier noise, and photomultipliers come into their own. The photomultiplier starts with a photocathode, just like the vacuum photodiode, but this is followed by an assembly of dynodes, each at a potential of about +100 V with respect to the last, and each emitting several secondary electrons for every one that hits it. The total gain of the multiplier depends on the number of stages and the applied voltage, but is typically 10^6 for 11 stages with a total applied voltage of 1 kV. The noise of a photomultiplier comes almost entirely from the photocathode dark current; the signal-to-noise ratio is normally degraded very little by the amplification. Dark currents are typically 0.1–1 nA, compared with a maximum output of around 100 μA, and can be reduced further by cooling. The detectivity is some two orders of magnitude above that of the silicon diode in Fig. 4.6. Standard photomultipliers have a time response of about 10 ns, and special tubes can be made about 10 times faster. Properly used, with the gain adjusted to avoid saturation, a photomultiplier can give a linear response over about nine orders of magnitude. At extremely low light levels (below about 10^5 photons s^{-1}) photomultipliers can be used in photon-counting mode – one pulse per detected photon.

The spectral response of photomultipliers is determined by the photocathode material and, where relevant, the window. Quantum efficiency peaks at about 25% in the visible region, falling, even for the most favourable materials, to 1% by 1 μm and falling also to about 10% in the ultra-violet. For ultra-violet applications useful discrimination against visible radiation is afforded by 'solar-blind' photocathodes which cut off at 300 or even 180 nm. At the short-

wavelength end the limit is set by the window material of the tube. LiF has the lowest limit, close to 100 nm; for shorter wavelengths the choice is between a phosphor-coated window and an open (windowless) tube.

Photo-ionization becomes a possible detection mechanism below about 130 nm, where the photon energy of >9 eV is sufficient to ionize gases. For continuous (d. c.) detection an ionization chamber is used, working on the plateau or saturation region of the curve of ion current versus voltage, since the ion current is then virtually independent of the chamber voltage and is proportional to the incident photon flux. The quantum efficiency, in terms of ion pairs per photon, may easily be 100% – indeed, if the photon energy is high enough for double ionization it may actually be greater than this. The pulsed version of the detector is essentially a Geiger counter: the original photoelectron from the incident photon is accelerated to make several collisions with gas molecules so that an avalanche takes place, using gas amplification. Both types are difficult to use in the wavelength region 104–30 nm because of the lack of transmitting windows. Below 30 nm cellulose and thin metal films start to transmit, and the detector can be used from there right on to the X-ray region. The filler gas at longer wavelengths is usually nitric oxide or a similar molecule, but the inert gases are preferable as soon as the photon energy reaches their respective ionization potentials. With a clever choice of window material and filler gas it is possible to arrange the long- and short-wavelength cut-offs to leave a fairly narrow band of sensitivity.

Photo-ionization detectors have also been used to measure absolute intensities and to calibrate sources as intensity standards in the vacuum ultraviolet. For 100% quantum efficiency the photon flux can be calculated from the output current of an ion chamber. The inert gases fulfil this requirement successively from 102 nm, the ionization limit of xenon, to 25 nm, where the ejected electrons have sufficient energy to cause secondary ionization even in helium so that the 1 : 1 correspondence breaks down. Photon counters can take over at this point, since they record separate pulses from each absorbed photon, irrespective of the number of secondary electrons.

4.13 Photographic emulsions

All the detectors so far considered give an output signal proportional to the incident radiant flux over a wide dynamic range. The response of a photographic emulsion to radiation is quite different. It is usually represented in the form shown in Fig. 4.7, in which density d is plotted against the logarithm of flux density (irradiance) I for constant exposure time t. d is defined as $\log_{10}(1/T)$, where T is the transmissivity of the blackened emulsion – i.e. the ratio of transmitted to incident intensity. The properties of the emulsion are characterized by three parameters, threshold, contrast and saturation, as shown on Fig. 4.7(a). Typical response curves for slow (insensitive), medium and fast

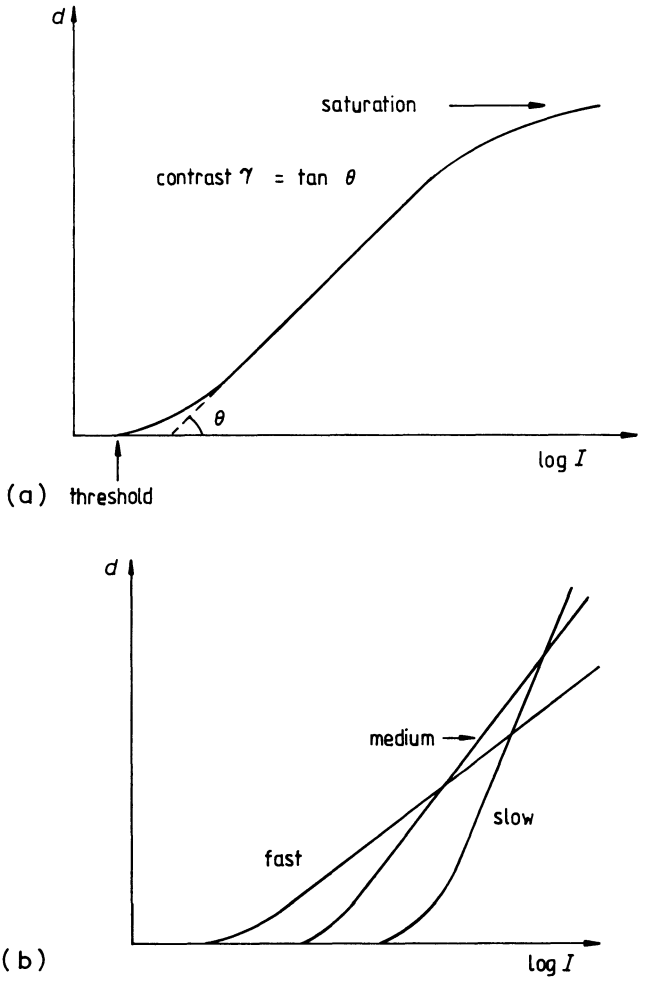

Figure 4.7 Photographic emulsion response curves: density d versus log I where $d=$ log(1/transmission) and I is incident flux density. The emulsion characteristics are defined in (a), and emulsions of different speed are compared in (b).

(sensitive) emulsions are given for comparison in Fig. 4.7(b). A fast emulsion means a low threshold, but as the contrast is usually low the film is not necessarily fast for long exposures.

Since a photographic emulsion is essentially an integrating device, it might be thought that d should depend on the product It, in which case Fig. 4.7 could equally well represent a plot of d versus log t. However, the silver halide grains are sensitized by a rather complicated multistage process, and the reciprocity law between irradiance and exposure time does not usually hold strictly.

Doubling the illumination and halving the time does not give the same blackening. d is actually a function of It^q, where q, the Schwarzchild factor, is a little less than unity.

Good quantitative intensity measurements are very hard to make photographically. The regions near threshold and saturation are obviously unsuitable, and the fairly linear section of the curve covers only about two orders of magnitude. Moreover, the slope of this region, the contrast γ, depends not only on the type of emulsion and the wavelength, but also on the conditions of exposure and development and even on the particular batch of film. The only safe procedure is to calibrate each individual film or plate with a set of known relative exposures obtained from a neutral (grey) wedge or step filter, a stepped slit or a rotating sector.

If neither accurate intensity measurements nor time resolution are required, the photographic emulsion has considerable assets. First, it is a multichannel detector, capable of recording a very large number of spectral elements simultaneously. It is not affected by fluctuations or drift of the source, unlike detectors that have to scan through the spectrum. Secondly, it gives better wavelength accuracy than a photoelectric detector in a high-resolution spectrograph; there are no problems from accuracy or reproducibility of scanning, and its resolution, limited by the grain size, is about 10 μm for a fine (slow) emulsion and 25 μm for a coarse (fast) one. Although the slit width for a photoelectric detector can in principle be as small as this, wider slits are usually required in practice for a sufficiently high signal level. Fourier transform spectroscopy is an exception to this generalization, as discussed in Chapter 7.

The wavelength range for normal photographic emulsions is about 230–700 nm. Special infra-red emulsions have been pushed up to 1200 nm, but they have very poor sensitivity. In the ultra-violet plates with emulsions almost free of gelatine are sensitive right down to the X-ray region, though the contrast is somewhat low.

4.14 Array detectors

Many scientific and technological activities require the conversion of an optical image to a set of electrical signals, and a detector combining the multichannel capability of photographic film with the linear response, large dynamic range and time resolution of a photoelectric detector is a most attractive prospect for spectroscopy. Until recently the size of the individual sensors was a serious limitation on spectral resolution, but the highest-resolution diode arrays are now close to photographic grain size, leaving the complexity of data handling and the cost as the two principal deterrents. The rate of development of these devices is such that details of their design and performance are likely to be out of date rather soon. We shall simply summarize the more important features of the devices currently available, making no attempt to discuss read-out and data-handling systems.

Two classes of array detectors are currently suitable for spectrometers. First, there are photodiode arrays, which are either one- or two-dimensional arrays of silicon p–n junctions similar to those described in section 4.12. They have the same spectral range, 1 μm to 200 nm, and quantum efficiency (60–70%), with dynamic range of 10^4–10^5. Linear arrays suitable for grating spectrometers may have up to 1024 sensors, spaced by as little as 25 μm (or even 12.5 μm). The two dimensional arrays required for an echelle grating or for spatial resolution are a little coarser. The actual resolution limit has to be taken as twice the sensor spacing. The arrays are normally run as charge-coupled devices (CCDs): each photodiode current is integrated on its own capacitance, and the stored charges are read off sequentially at a cycle rate of anything up to a few MHz.

The second class is related to the photomultiplier rather than the photodiode. Channel multipliers are photomultiplier tubes in which the multiplication is achieved down a single curved channel coated with a thin resistive layer, instead of along a discrete dynode chain. These channels can be miniaturized to form a microchannel array with individual channel (pixel) sizes down to 25×25 μm. As in a conventional photomultiplier tube, the spectral range is set by the photocathode and window materials anywhere from the visible to the soft X-ray region. Quantum efficiencies, at around 5–15%, and multiplication factors, $\sim 10^6$, are as described in section 4.12.

References

The order is alphabetical within each group, and within each sub-group in the last group.

GENERAL: RADIOMETRY, BLACKBODY, EINSTEIN COEFFICIENTS, ETC.

*1. Corney, A. (1977) *Atomic and Laser Spectroscopy*, Oxford University Press.
*2. Ditchburn, R.W. (1963) *Light*, Blackie, London.
3. Grun, F. and Becherer, R.J. (1979) *Optical Radiation Measurements*, Vol. 1, Academic Press, New York.
*4. Loudon, R. (1977) *Quantum Theory of Light*, Oxford University Press.

SOURCES AND DETECTORS, GENERAL

5. Harrison, G.R., Lord, R.C. and Loofbourow, J.R. (1948) *Practical Spectroscopy*, Prentice-Hall, Englewood Cliffs, New Jersey.
6. Keyes, R.J. (ed.) (1976) *Optical and Infrared Detectors* (*Topics in Applied Physics*, Vol. 19), Springer-Verlag, Berlin.
7. Kingston, R.H. (1978) *Detection of Optical and Infrared Radiation*, Springer-Verlag, Berlin.
8. Sawyer, R.A. (1961) *Experimental Spectroscopy*, Dover, New York.

* Further reading, not cited in text.

INFRA-RED REGION

9. Chantry, G.W. (1984) *Long-wave Optics* (2 Vols), Academic Press, New York.
10. Conn, G.K.T. and Avery, D.G. (1960) *Infrared Methods*, Academic Press, New York.
11. Kimmit, M.F. (1970) *Far Infrared Techniques*, Pion, London.
12. Kruse, P.W., McGlauchlin, L.D. and McQuistan, R.B. (1962) *Elements of Infra-red Technology*, Wiley, New York.
13. Martin, A.E. (1966) *Infra-red Instrumentation and Techniques*, Elsevier, Amsterdam.
14. Martin, D.H. (ed.) (1967) *Spectroscopic Techniques for Far Infra-red, Sub-millimetre and Millimetre Waves*, North-Holland, Amsterdam.
15. Robinson, L.C. (1973) Physical principles of far infra-red radiation, in *Methods of Experimental Physics*, Vol. 10 (ed. L. Marton), Academic Press, New York.
16. Smith, R.A., Jones, F.E. and Chasmar, R.P. (1968) *Detection and Measurement of Infra-red Radiation*, Oxford University Press.

ULTRA-VIOLET REGION

17. Garton, W.R.S. (1966) Spectroscopy in the vacuum ultra-violet, in *Advances in Atomic and Molecular Physics*, Vol. 2, Academic Press, New York.
18. Samson, J.A.R. (1967) *Techniques of Vacuum Ultra-violet Spectroscopy*, Wiley, New York.
19. Zaidel, A.N. and Schreider, E.Y.A. (1970) *Vacuum Ultraviolet Spectroscopy*, Humphrey Science, Ann Arbor, Michigan.

ULTRA-VIOLET INTENSITY STANDARDS

20. Key, P.J. and Preston, R.C. (1977) The 'blackbody arc' *Appl. Opt.*, **16**, 2477.
21. The deuterium lamp: (NBS) Saunders, R.D., Ott, W.R. and Bridges, J.M. (1978) *Appl. Opt.*, **17**, 593; (NPL) Key, P.J. and Preston, R.C. (1980) *J. Phys. E*, **13**, 866.
22. The argon mini-arc: (NBS) Bridges, J.M. and Ott, W.R. (1977) *Appl. Opt.*, **16**, 367. (NPL) Preston, R.C., Brookes, C. and Clutterbuck, F.W.J. (1980) *J. Phys. E*, **13**, 1206.

SPECIAL SOURCES

Information on standard laboratory sources can be found in [5–19] above. The references below give further information on sources that are too new or too rarely used to be covered adequately or at all in [5–19].

Flames

Gaydon, A.G. and Wolfhard, H.G. (1960) *Flames*, Chapman and Hall, London.

Arcs

Foster, E.W. (1964) Measurement of oscillator strengths, *Rep. Prog. Phys.*, **27**, 469.

Wiese, W.L. (1962) Electric arcs, in *Methods of Experimental Physics*, Vol. 7B (ed. L. Marton), Academic Press, New York.

Hollow cathodes

Caroli, S. (ed.) (1985) *Improved Hollow Cathode Lamps for Spectroscopy*, Horwood, Chichester.

Tolansky, S. (1947) *High Resolution Spectroscopy*, Methuen, London.

Shock-tubes

Kolb, A.C. and Griem, H.R. (1962) High temperature shock waves, in *Atomic and Molecular Processes* (ed. D.R. Bates), Academic Press, New York.

Pain, H.J. and Rogers, E.W.E. (1962) Shock waves in gases, *Rep. Prog. Phys.*, **25**, 287.

Plasma sources

Burgess, D. D. (1972) Spectroscopy of laboratory plasmas, *Space Sci. Rev.*, **13**, 493.

Fawcett, B.C. (1974) Recent Progress in the Classification of the Spectra of Highly Ionised Atoms in *Advances in Atomic and Molecular Physics* Vol. 10 (eds D.R. Bates and B. Bederson) p. 223, Academic Press, New York.

Glasstone, S. and Lovberg, R.H. (1960) *Controlled Thermonuclear Reactions*, Van Nostrand, New York.

Green, T.S. (1963) *Thermonuclear Power*, Newnes, London.

Gross, R.A. and Miller, B. (1970) Plasma heating by strong shock waves, in *Methods of Experimental Physics*, Vol. 9A (eds H.R. Griem and R.H. Lovberg), Academic Press, New York.

Mather, J.W. (1970) Dense plasma focus, *ibid.*, Vol. 9B.

Niblett, G. B. F. (1970) Production and containment of high density plasmas, in *Physics of Hot Plasmas* (eds B.J. Rye and J.G. Taylor), Oliver and Boyd, Edinburgh.

Beam foil

Bashkin, S. (1968) Beam foil spectroscopy, *Appl. Opt.*, **7**, 2341.

Bashkin, S. (ed.) (1976) *Beam Foil Spectroscopy* (*Topics in Current Physics*, Vol. 1), Springer-Verlag, Berlin.

Inductively coupled plasma

Thompson, M. and Walsh, J.N. (1983) *A Handbook of Inductively Coupled Plasma Spectrometry*, Blackie, London.

Boumans, P.W.J.M. (1987) *Inductively Coupled Plasma Emission Spectroscopy*, Wiley, Chichester.

Synchrotron radiation

Godwin, R.P. (1969) Synchrotron radiation as a light source, *Springer Tracts in Modern Physics*, Vol. 51, Springer-Verlag, Berlin.

Laser plasma continuum

Carroll, P.K., Kennedy, E.T. and O'Sullivan, G. (1980) Laser-produced continua for absorption spectroscopy in the vacuum ultraviolet and XUV, *Appl. Opt.*, **19**, 1454.

Kuhne, M. (1982) Radiometric comparison of a laser-produced plasma and a BRV-source plasma at normal incidence, *Appl. Opt.*, **21**, 2124.

In addition, up-to-date information on both detectors and 'conventional' sources is to be found in manufacturers' catalogues.

FIVE

Dispersion and resolving power: prism spectrographs

5.1 General remarks

Dispersion can be achieved by refraction, diffraction or interference – that is, by prisms, gratings or interferometers. This distinction is in a way rather artificial, because all these instruments work by superposing a number of rays of varying phase, the number being infinite for the prism, 10^5–10^3 for the grating and 50–2 for the interferometer. The important properties of a disperser are its spectral range, its resolving power and dispersion, its light grasp and its ambiguity or ease of interpretation. On spectral range gratings win easily, as they can be used all the way from X-rays to microwaves. Prisms require a transparent solid, which limits the range to between 40 μm and 120 nm. Interferometers require a beam-splitter; in practice this imposes the same limit at the short-wavelength end, but thin films or grids can be used above the crystal transmission limit at long wavelengths. In respect of maximum resolving power, the order of merit is interferometer, grating, prism, with an order of magnitude between each. Light grasp is inversely related to resolving power, but a comparison of light grasp for the same resolving power favours the interferometer by one to two orders of magnitude. In ease of interpretation the order is reversed: only prisms give directly unambiguous spectra. Overlapping orders may occur with gratings and always need attention with Fabry–Perot interferometers; the Michelson interferometer used in Fourier transform spectroscopy can indeed give an unambiguous spectrum, but only after the output has been Fourier-transformed.

We revert to the comparison of interferometers with diffraction grating instruments at the end of Chapter 7 after both types have been described. In this chapter a general discussion of resolving power applicable to all three types of spectrometer is followed by some remarks on slit width and illumination relevant to prisms and gratings. The description of prism instruments that follows has been kept brief because their limited spectral range and resolving power renders them relatively unimportant for most of the line shape and intensity measurements with which the last part of this book is concerned.

5.2 Dispersion and resolving power

Leaving aside for the moment Fourier transform spectrometers, dispersing instruments separate different wavelengths by spreading them out spatially. Dispersion is the measure of this spreading. If rays of wavelengths λ and $\lambda+\mathrm{d}\lambda$ emerge from the disperser at angles θ and $\theta+\mathrm{d}\theta$, respectively, the *angular dispersion* is defined as

$$\text{angular dispersion}=\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} \quad \text{rad nm}^{-1} \tag{5.1}$$

The *linear dispersion* is the differential linear separation of the two wavelength peaks in the focal plane of the camera lens or concave grating, $\mathrm{d}l/\mathrm{d}\lambda$. In practice more use is made of the reciprocal dispersion $\mathrm{d}\lambda/\mathrm{d}l$. The correct SI unit for this is nm mm^{-1}, but the traditional form Å mm^{-1} makes the meaning considerably clearer. For a focal length of f mm,

$$\frac{\mathrm{d}\lambda}{\mathrm{d}l}=\frac{1}{f}\frac{\mathrm{d}\lambda}{\mathrm{d}\theta} \quad \text{nm mm}^{-1} \quad \text{or} \quad \text{Å mm}^{-1} \tag{5.2}$$

Explicit expressions for the dispersion of prisms, gratings and Fabry–Perot interferometers will be derived when the individual instruments are described in this and the next two chapters.

Resolving power is a more fundamental concept. A spectral 'line' produced by an ideal spectrometer illuminated by an ideal source of monochromatic light has a finite width set by diffraction or bandwidth limits. The intensity distribution of output from a monochromatic input is known as the *instrument function*, and we shall assume for the moment an ideal instrument function, as produced by a spectrometer that is free of optical aberrations and mechanical defects. The smallest measurable wavelength difference $\delta\lambda$ is related to the width of this function. More precisely, if λ and $\lambda+\delta\lambda$ are the wavelengths of two monochromatic lines that are just resolved by the spectrometer, the *resolving power* is defined as $\lambda/\delta\lambda$. It is obviously a dimensionless quantity, and can equally well be applied to wavenumber or frequency difference, i.e.

$$\mathscr{R}=\frac{\lambda}{\delta\lambda}=\frac{\sigma}{\delta\sigma}=\frac{\nu}{\delta\nu} \tag{5.3}$$

What is meant by 'just resolved' depends to some extent on the shape of the instrument function, but it is possible, as will now be shown, to derive results applicable to any instrument within a factor of two. It should be emphasized that resolving power cannot be increased simply by increasing dispersion – for example, by using a larger value of f in equation 5.2 – since the instrumental width of each line is thereby increased in exact proportion to the line separation.

Ultimate limits on resolving power are most easily understood by considering first a two-beam interferometer in which an optical path difference x is imposed between the beams. An interference maximum occurs for wavelength λ whenever

$x = m\lambda$, with m an integer. To distinguish between λ_1 and λ_2 we might reasonably require to be able to count $m \pm 1$ maxima of λ_2 for m of λ_1; the resolution criterion is therefore that x must reach a value L such that

$$L = m\lambda_1 = (m-1)\lambda_2 = (m-1)(\lambda_1 + \delta\lambda)$$

Assuming $\delta\lambda \ll \lambda_1$, this becomes

$$\frac{\lambda}{\delta\lambda} = m \quad \text{or} \quad \mathscr{R} = m = L/\lambda \tag{5.4}$$

This suggests a basic definition of $\mathscr{R}$ as the number of wavelengths in the extreme path difference.

It is often useful to think in terms of *resolution limit*, which is the interval $\delta\lambda$ (nm) between two lines that are just resolved; it may equally well be expressed in wavenumbers $\delta\sigma$ (cm^{-1}) or frequency $\delta\nu$ (Hz). From equations 5.3 and 5.4 we have the convenient expression

$$\delta\sigma = 1/L \quad \mathrm{cm}^{-1} \tag{5.5}$$

This emphasizes the fundamental limitation on wavelength measurement, for it is just one way of writing the Fourier bandwidth theorem:

$$\delta\nu\,\delta t \sim 1 \quad \text{or} \quad \frac{\delta\nu}{c} c\,\delta t = \delta\sigma L \sim 1 \tag{5.6}$$

where δt is the maximum time delay imposed by the interferometer.

Prisms and gratings deviate the incident wavefront and thereby impose a wavelength-dependent path difference $L(\lambda)$ across it as shown in Fig. 5.1. For a grating the path difference between adjacent rulings must be an integral number of wavelengths, $m\lambda$; so $L = Nm\lambda$, where N is the number of rulings. Again, we require for resolution that L_1 and L_2 should differ by a wavelength (amounting to an extreme phase difference of 2π between the two diffracted wavefronts):

$$\delta L = \lambda \quad \text{where} \quad \delta L = Nm\delta\lambda \tag{5.7}$$

therefore

$$Nm\delta\lambda = \lambda \quad \text{and} \quad \mathscr{R} = Nm = L/\lambda \tag{5.8}$$

This not only reproduces the interferometer definition (equation 5.4), but also shows that it is identical with the better-known way of expressing the resolving power of a grating as (order) × (number of rulings).

The resolving power of a prism or grating is conventionally derived from the diffraction-limited instrument function. It is worth working through this derivation to show explicitly that it leads to the same result. Figure 5.2(a) shows schematically Fraunhofer diffraction at an aperture of width D, which represents the projected width of the prism or grating (Fig. 5.2(b)). The grating rulings and the slit length are both perpendicular to the plane of the figure, and the slit width is assumed for the present purpose to be negligibly small. The Fraunhofer

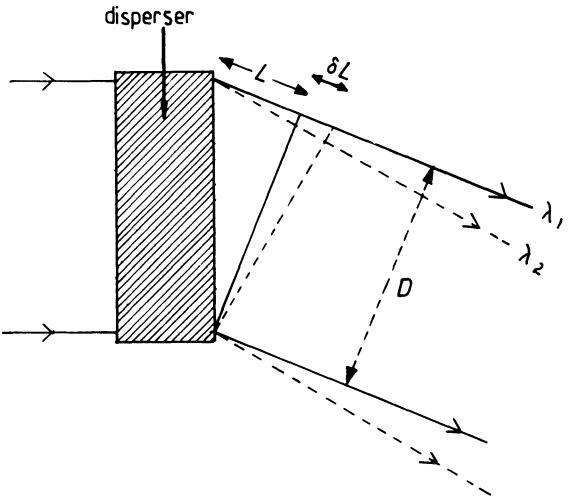

Figure 5.1 Resolution criterion for a prism or grating. The path difference L imposed across the beam of width D is a function of wavelength, and two wavelengths λ_1 and λ_2 are taken to be resolved if δL is one wavelength.

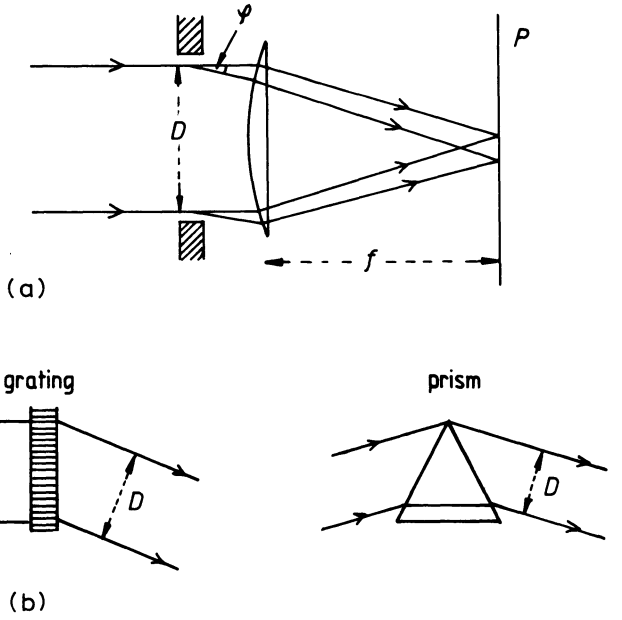

Figure 5.2 (a) Ray diagram for Fraunhofer diffraction at a rectangular aperture of width D. The angular spread φ is much exaggerated. (b) Effective aperture D for a grating and a prism.

diffraction intensity distribution is derived in standard optics texts (e.g. [1, 3, 5]). In terms of the angular spread φ it is given by

$$I = I_0\left(\frac{\sin\alpha}{\alpha}\right)^2 \equiv I_0 \operatorname{sinc}^2\alpha \quad \text{with} \quad \alpha = \frac{\pi D\varphi}{\lambda} \tag{5.9}$$

where I_0 is the intensity at the central maximum ($\alpha = \varphi = 0$) and the spread is assumed small enough for $\sin\varphi = \varphi$. The first minimum on either side occurs for $\alpha = \pm\pi$, or $\varphi_m = \pm\lambda/D$, as shown in Fig. 5.3.

Suppose we now add to the picture the diffraction pattern for a second line of wavelength $\lambda + \delta\lambda$. The central maxima of the two lines are separated by an angle $\delta\theta$ that depends on the angular dispersion according to $\delta\theta = (\mathrm{d}\theta/\mathrm{d}\lambda)\delta\lambda$. If $\delta\theta$ is appreciably smaller than φ_m the two lines will not be distinguishable. The condition for resolution normally adopted is Rayleigh's criterion, which requires that the central maximum of one line should fall on the first minimum of the other, so that $\delta\theta = \varphi_m$ as shown in Fig. 5.4. Therefore,

$$\mathscr{R} = \frac{\lambda}{\delta\lambda} = \frac{\lambda}{\varphi_m}\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = D\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} \tag{5.10}$$

It is easily shown that this definition is identical with that of equation 5.8. If W is the width of the dispersing element in Fig. 5.1, we have $L = W\sin\theta$

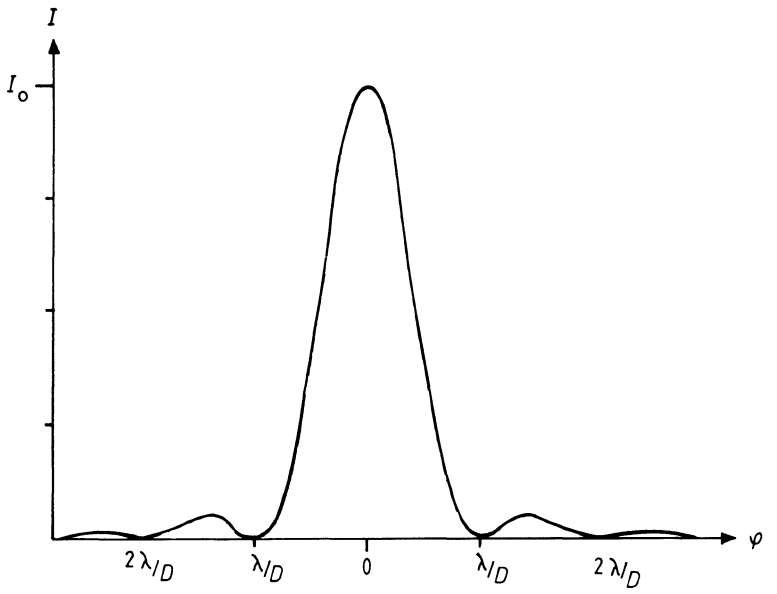

Figure 5.3 Fraunhofer diffraction intensity distribution: $I = I_0 \operatorname{sinc}^2(\pi D\varphi/\lambda)$.

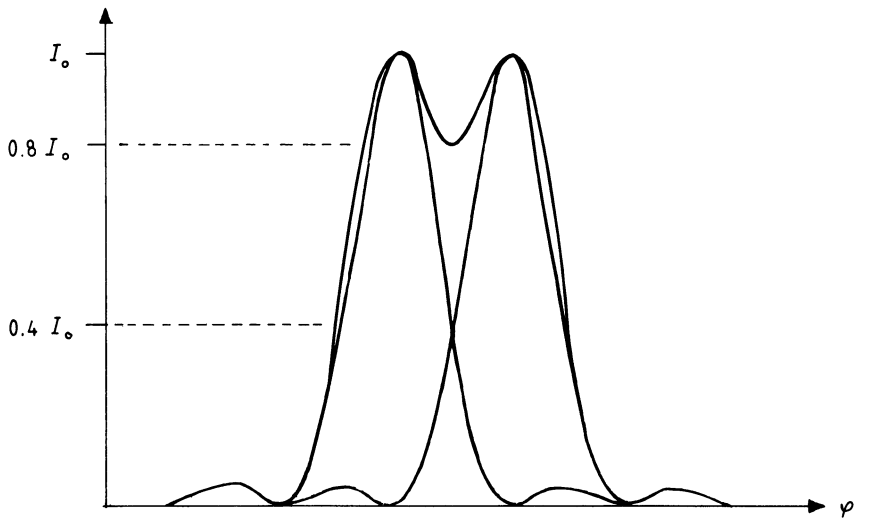

Figure 5.4 Rayleigh's criterion for resolution. The maximum of each line falls on the first minimum of the other. The heavy curve is the superposition of the two individual curves and has a 20% dip.

and $D = W\cos\theta$. Therefore,

$$\delta L = \left(\frac{dL}{d\lambda}\right)\delta\lambda = \frac{dL}{d\theta}\frac{d\theta}{d\lambda}\delta\lambda = W\cos\theta\frac{d\theta}{d\lambda}\delta\lambda = D\frac{d\theta}{d\lambda}\delta\lambda$$

This satisfies equation 5.10 if δL is set equal to λ, which is exactly the requirement of equation 5.7.

For an instrument function that does not have a diffraction intensity distribution, Rayleigh's criterion is adapted to state that the dip between the peaks of two equal lines must be 80% of the peak of either. This is the same dip–peak ratio as in the diffraction case, since the sinc^2 function falls to 0.4 of its maximum half-way between the centre and the first minimum (Fig. 5.4)

In practice any resolution criterion is somewhat arbitrary. Lines considerably closer than the resolution limit may be distinguished if their intensity is similar and the signal-to-noise raito is good. Conversely, a weak line in the wing of a strong one will not necessarily be noticeable, even at the resolution limit, in a noisy spectrum.

5.3 Slit width

The remarks on slit width in this section apply only to prism and grating spectrometers. Fabry–Perot and Michelson interferometers have circular

symmetry about the optic axis, and different restrictions on input aperture apply, as is seen in Chapter 7.

The instrument function of a prism or grating spectrometer is the image of the slit in light of the appropriate wavelength. A wide line is obtained from a wide slit, but narrowing the slit does not decrease the linewidth beyond the diffraction limit. As shown in the previous section, the diffraction angular spread, central maximum to first minimum, is λ/D, and the angle λ/D is also the angular separation of the points in the image plane at which the intensity drops to 0.4 of its peak value. We can therefore take $f\lambda/D$ as (closely enough) the full-width half-maximum (FWHM) of the diffraction-limited line, where f is the focal length of the camera lens. ('Camera lens' is a slightly anachronistic but convenient term for the imaging element of a spectrometer; it may often stand for a mirror, and it does not imply that the instrument is necessarily photographic.) The instrument function from a slit of finite width w is the convolution of its image w' with the diffraction profile. For simplicity we assume the collimating and camera lenses to have equal focal lengths, so $w'=w$. The FWHM of the convolution, w_c, is close to $f\lambda/D$ when $w<f\lambda/D$, and it approaches w when $w>f\lambda/D$, as shown by the solid curve in Fig. 5.5(a). The optimum slit width w_0 has to be a compromise between a narrow instrument function w_c and a high light throughput, since the latter is proportional to w. The broken curve in Fig. 5.5(a) shows how the flux density I_0 at the peak of the distribution varies with w. The optimum is taken to be

$$w_0=f\lambda/D \tag{5.11}$$

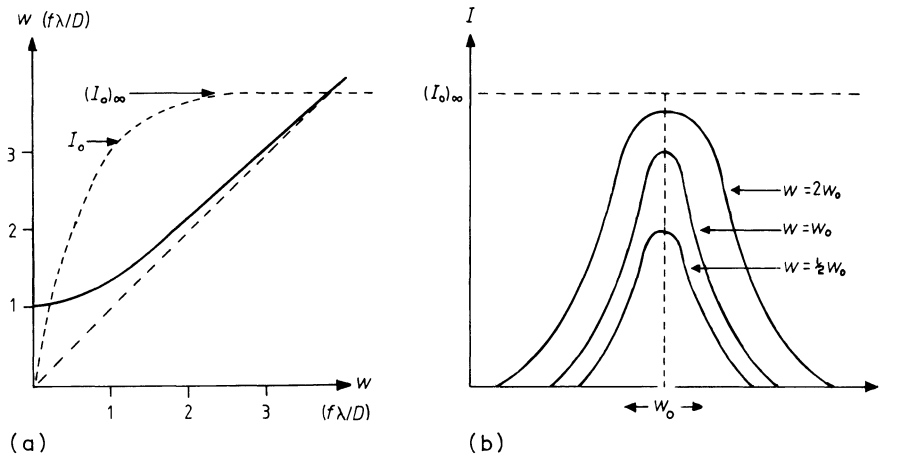

Figure 5.5 Effect of slit width w on (a) image width w_c and peak flux density I_0 and (b) intensity distribution in image plane. In (a) both w and w_c are in units of $f\lambda/D$, and w_c is the width of the convolution at $0.4\times$ maximum. (b) illustrates the justification for taking $w_0(=f\lambda/D)$ as the optimum compromise between intensity and width.

since smaller values of w sacrifice flux for every little reduction in w_0. This is also evident from Fig. 5.5(b), which shows the intensity distribution for $w=\frac{1}{2}w_0$, $w=w_0$ and $w=2w_0$.

These arguments apply strictly only to non-coherent illumination of the slit, which is the case when a thermal source or an image of it is fairly close to the slit. The effects of coherent illumination, as with a collimated beam or a laser, are a little different; a fuller discussion may be found in [2, 4].

In most actual spectrometers f/D is in the range 10–40, and the optimum slit width from the near infra-red to the near ultra-violet is of order 10 μm. Narrower slits present mechanical problems, and 10 μm is also the resolution limit of fine-grain photographic emulsions. With photoelectric detectors the optimum slit width is frequently exceeded, particularly in the infra-red, to provide sufficient light flux.

In the visible region the slit may easily be adjusted by looking at the diffraction pattern formed near the prism or grating when the source is distant enough for the slit itself to act as a diffracting aperture coherently illuminated. The diffraction pattern, as imaged by the collimating lens of focal length f' has a FWHM of $f'\lambda/w$, and this just fills the grating width D when $D\simeq f'\lambda/w$: i.e. $w=w_0$ if $f=f'$.

It needs to be remembered that the blackening of a photographic emulsion depends on the flux density, and for a line whose intrinsic width is less than w_0 this is not increased by widening the slit appreciably beyond w_0: the additional flux simply goes into blackening more grains. For wide lines and continuous spectra the flux density *does* increase with w. Photoelectric detectors integrate over the area of the exit slit of the spectrometer, and the signal must always increase with w provided the entrance and exit slits are widened together. The additional signal is of course acquired at the expense of resolution.

5.4 Throughput and illumination

The greatest power that can be put into a spectrometer is determined by the product of the radiance of the source and the throughput of the instrument. Throughput was defined in equation 4.5 as the product of cross-sectional area and solid angle, a quantity that is invariant through an optical system. An obvious place to measure throughput is at the entrance to the spectrometer, where it is equal to the product of the area of the entrance aperture and the solid angle subtended there by the disperser or the interferometer plates. To simplify the discussion it is assumed for the moment that we are concerned with a grating spectrometer, with a slit as the entrance aperture. The important gain in throughput afforded by an interferometer with a circular input aperture is emphasized in Chapter 7; the remarks on illumination that follow in this section are equally applicable to either system.

The input solid angle is determined by the focal ratio D/f of the collimator, assuming the diameter D of this to be matched to the projected diameter of the

prism or grating, which is of course the most expensive optical element. For a slit of width w and height h,

$$\text{throughput } T = wh(1/f)^2 \pi(D/2)^2 = wh\frac{\pi}{4}\left(\frac{D}{f}\right)^2$$

To take full advantage of this throughput the source must be either close enough to the slit or large enough to subtend at the slit a solid angle at least as large as that of the collimator. This requires a source of width AA′ and height BB′, as shown in Fig. 5.6. Every point between A and A′ in one dimension and aa′ in the other sends light through the whole collimator; the regions aB and a′B′ illuminate only part of the collimator but nevertheless contribute to the total throughput. If the collimator is properly filled in this way, the total power W reaching the detector is given by

$$W = LT\tau = Lwh\frac{\pi}{4}\left(\frac{D}{f}\right)^2\tau \tag{5.12}$$

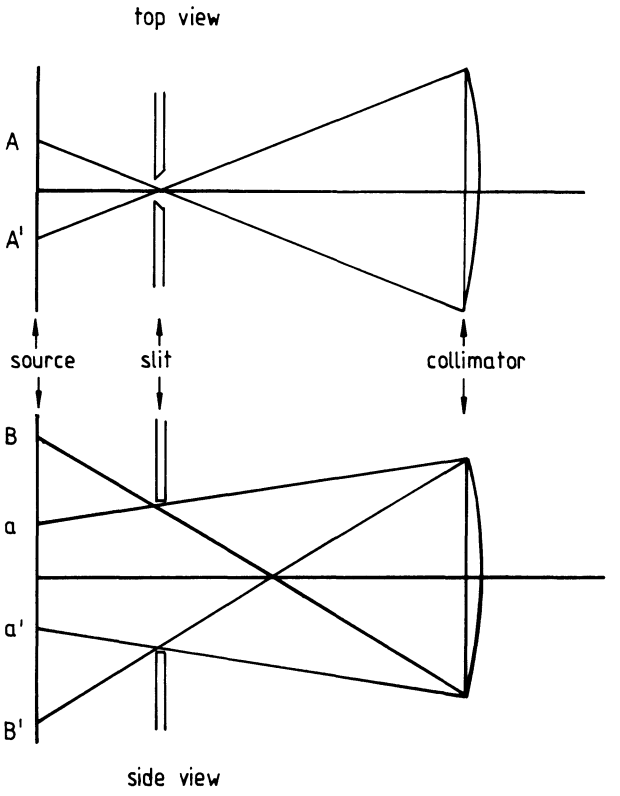

Figure 5.6 Illumination of collimator lens. For maximum illumination the source should have width AA′ and height BB′.

where L is the radiance of the source and τ is the optical efficiency of the spectrometer, incorporating all transmission and reflection losses.

It is important to be clear about the purpose of a condensing lens between the source and the spectrometer. If the geometry of Fig. 5.6 is satisfied, W cannot be increased by any condensing system. This is a consequence of the invariance of L referred to in section 4.1. At first sight it appears that one ought to gain by focusing a reduced image of a large source on the slit. Certainly this looks 'brighter', but since the radiance is unchanged the flux is spread through a larger solid angle and some of it misses the collimator, as shown in Fig. 5.7(a). This 'wasted' flux is actually worse than useless, because it adds to the scattered light background. The ideal condensing system is shown in Fig. 5.7(b): the condensing lens is stopped down to match the solid angles on either side of the slit. If the source is actually very small (a capillary tube, for example) it may be desirable to form an enlarged image on the slit so as to use the full slit height, as shown in Fig. 5.7(c); the condenser should again be stopped down to match the solid angles. The alignment of source and condenser is very critical in a properly matched system: the source need be only 2° off the optic axis of an $f/25$ spectrometer for the entire beam to miss the collimator.

If any actual experiment there may be other important considerations in the illumination of the slit. If uniform illumination is required it is better *not* to focus the source on the slit. Conversely, if spatial resolution is wanted, or if only a part of the source is of interest, it is essential not only to image the source on the slit, but also to remember that condensing lenses are usually uncorrected for chromatic aberration – the ultra-violet image of the source may be several centimetres outside the slit if the focusing is done visually.

5.5 Prism instruments

A conventional prism spectrograph is shown in Fig. 5.8. Light from the slit S is collimated by the lens L_1, refracted by the prism, and focused by the camera lens L_2 on the photographic plate P, where the spectral lines can be thought of as images of the slit in light of different wavelengths. For visual observation, L_2 is the objective of a telescope having its eyepiece focused on the plane P. With photoelectric detection the spectrum is usually scanned by tracking the exit slit and detector together along the plane P. This arrangement is the most usual one for small or medium-sized instruments not required to act as monochromators. The two lenses generally have the same focal length, giving unit magnification, and a single 60° prism is used.

Modifications of this basic arrangement are desirable for different applications. We survey here the most important ones; fuller details may be found in [2, 4, 6].

For large instruments, with a focal length of 1 m or more, the straight-through arrangement makes for an awkward shape, and the Littrow mounting usually

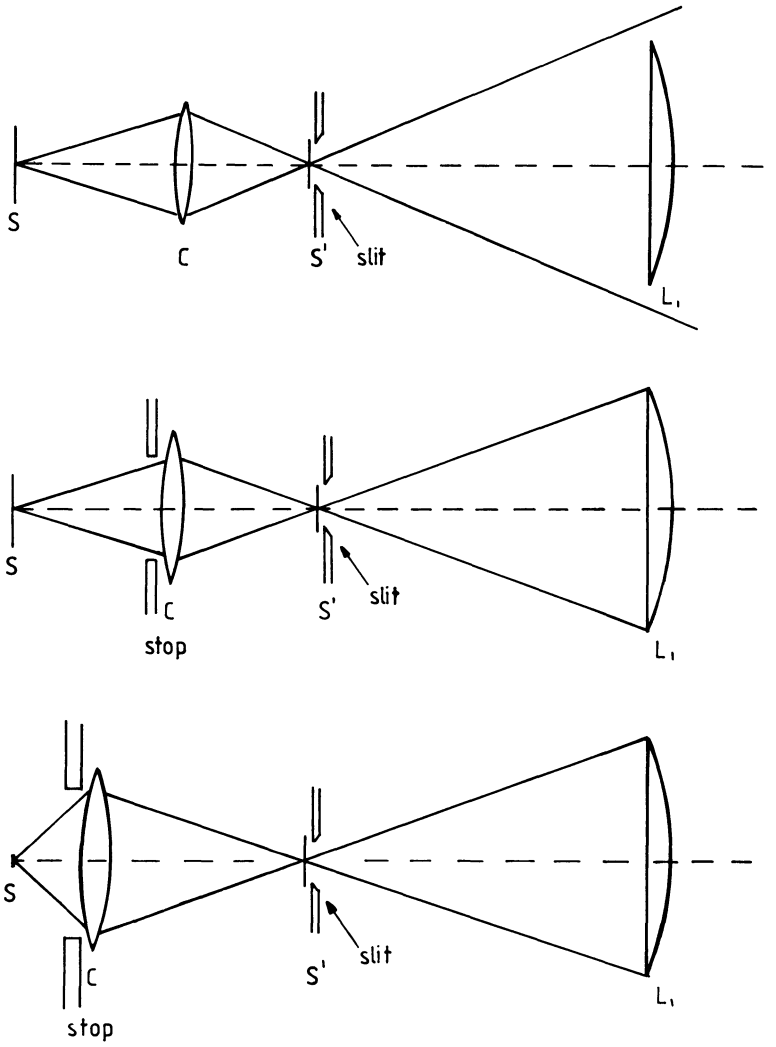

Figure 5.7 Matching of condensing system to spectrometer aperture. S′ is the image of the source S formed on the slit by the condensing lens C, and L_1 is the collimating lens of the spectrometer. In (a) the collimator is overfilled and in (b) a stop is used to match the solid angles. (c) shows the matched arrangement for a small source.

replaces it. This uses a 30° prism with a silvered back or with a mirror immediately behind it, so that the light retraces its path through the instrument. The same lens serves as both collimating and camera lens, saving expense as well

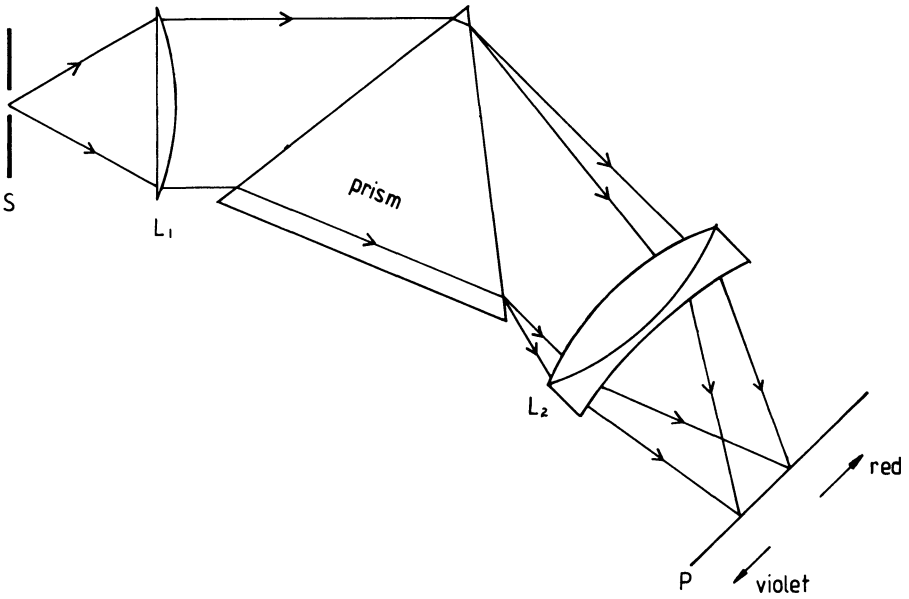

Figure 5.8 Components of a conventional prism spectrometer. S is the slit, L_1 and L_2 the collimating and camera lenses, and P the focal plane of L_2.

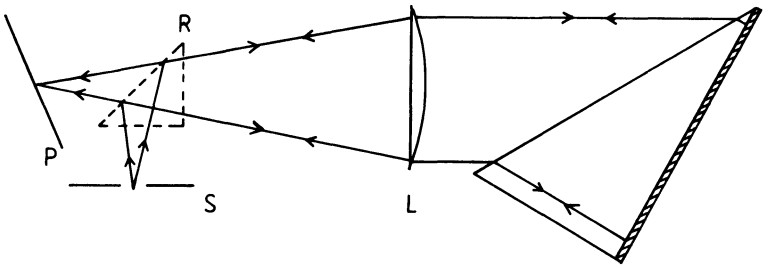

Figure 5.9 Littrow mounting for prism spectrometer, using a 30° back-reflecting prism. L acts as both collimating and camera lens. The slit S and reflecting prism R are above the plane of the rest of the diagram.

as space. The light is usually brought in from the side with the help of a totally reflecting prism, as shown in Fig. 5.9. A disadvantage of this type of mounting is that scattered light from reflections at the lens surfaces goes straight back to P.

For a monochromator it is important to have the emergent beam at a fixed angle, whatever the wavelength. The simplest constant deviation mounting is that originally devised by Wadsworth, in which a 60° prism is combined with a plane mirror as shown in Fig. 5.10. The same effect – a constant deviation of 90° – can be achieved with a four-sided prism, which effectively replaces the

mirror by a total internal reflection. Another method, at the cost of reduced dispersion, is to combine three 60° prisms of two different types of glass. This is the form usually adopted for small direct-vision spectroscopes.

When throughput is of paramount importance, as in stellar or Raman spectroscopy, it becomes essential to use an instrument of very wide aperture. Whereas most standard spectrometers have focal ratios between *f*/10 and *f*/25, very fast instruments may have *f*/1, or even *f*/0.5. The problems are those of lens design and are outside the scope of this discussion.

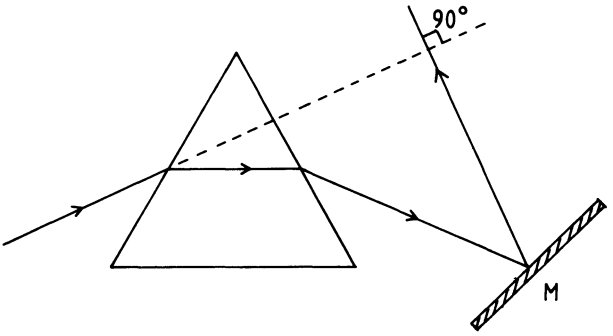

Figure 5.10 Wadsworth (constant deviation) mounting. The mirror M rotates with the prism so that the deviation is always 90°.

5.6 Deviation and dispersion

Figure 5.11 shows the path of a ray through a principal section of a prism of angle α. It can be seen from the geometry that the deviation θ is given by

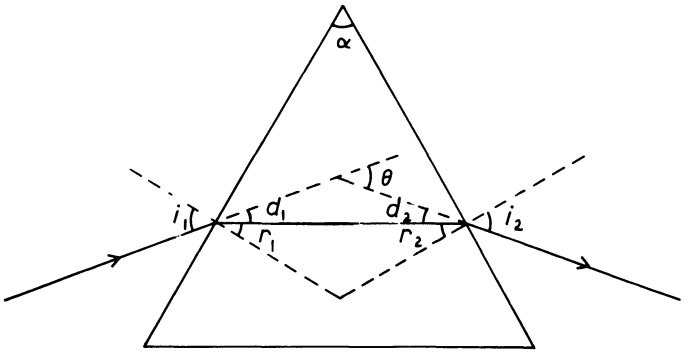

Figure 5.11 Path of ray through principal section of prism. The deviation θ is the sum of the deviations d_1 and d_2 at the two faces. i_1, r_1 and i_2, r_2 are the angles of incidence and refraction, respectively, at the two faces.

$\theta = d_1 + d_2$, where $d_1 = i_1 - r_1$ and $d_2 = i_2 - r_2$. But $\alpha = r_1 + r_2$, so $\theta = i_1 + i_2 - \alpha$. Prism instruments are normally used near the position of minimum deviation, for which it can be shown that the ray traverses the prism symmetrically. i_1 is then equal to i_2, and the ray passes through the prism parallel to the base, so that $\theta = 2i - \alpha$ and $\alpha = 2r$. Combining these relations with Snell's law of refraction, $\sin i = n \sin r$:

$$\sin i = \sin\left(\frac{\theta+\alpha}{2}\right) = n \sin\left(\frac{\alpha}{2}\right) \tag{5.13}$$

The change in deviation θ with wavelength determines the angular dispersion of the prism. Writing

$$\frac{d\theta}{d\lambda} = \frac{d\theta}{dn}\frac{dn}{d\lambda}$$

it can be seen from equation 5.13 that the angular dispersion depends on the prism angle α and on the dispersion of the prism material, $dn/d\lambda$. Differentiating equation 5.13,

$$\frac{d\theta}{dn} = \frac{2\sin(\alpha/2)}{\cos\frac{1}{2}(\theta+\alpha)} = \frac{2\sin(\alpha/2)}{[1-\sin^2\frac{1}{2}(\theta+\alpha)]^{1/2}} = \frac{2\sin(\alpha/2)}{[1-n^2\sin^2(\alpha/2)]^{\frac{1}{2}}}$$

Therefore

$$\frac{d\theta}{d\lambda} = \frac{2\sin(\alpha/2)}{[1-n^2\sin^2(\alpha/2)]^{1/2}}\frac{dn}{d\lambda} \tag{5.14}$$

The dispersion obviously increases with increasing α, but so too does the amount of material required for a prism of given aperture. An angle of 60° is usually taken as the best compromise. Putting $\sin(\alpha/2) = \frac{1}{2}$, equation 5.14 becomes

$$\frac{d\theta}{d\lambda} = \frac{1}{[1-(n/2)^2]^{1/2}}\frac{dn}{d\lambda} \tag{5.15}$$

It so happens that for values of n in the range 1.4–1.6 – which includes most suitable glasses and crystals – the first term on the right-hand side is equal to n to within 4% (the relation is exact for $n = \sqrt{2}$). To a good approximation, therefore,

$$\frac{d\theta}{d\lambda} \simeq n\frac{dn}{d\lambda} \tag{5.15a}$$

The dispersion of any material is a function of wavelength and increases sharply near a region of absorption. Prisms therefore disperse most efficiently near their absorption limits. A few figures for refractive index and dispersion in the visible and ultra-violet are given in Table 5.1, together with the angular dispersion of a 60° prism of the material. The figures for glass should be taken as

Table 5.1 Refractive index and dispersion of prism materials. The last three columns refer to a 60° prism

λ (nm)	Material	n	$dn/d\lambda$ ($\Delta n \times 10^3$ per nm)	$d\theta/d\lambda$ (mrad nm^{-1})	$dl/d\lambda$ (mm nm^{-1}) (f=0.5 m)	$d\lambda/dl$ (nm mm^{-1}) (f=0.5 m)
200	Fused silica	1.55	1.3	2.1	1.0	1.0
	Crystal quartz	1.65	1.6	2.8	1.4	0.7
300	Fused silica	1.49	0.28	0.43	0.21	4.6
	Crystal quartz	1.58	0.32	0.53	0.26	3.8
400	Fused silica	1.47	0.11	0.16	0.08	12.5
	Crown glass	1.53	0.13	0.20	0.10	10.0
	Flint glass	1.76	0.51	1.06	0.53	1.9
600	Crown glass	1.52	0.036	0.054	0.03	37.0
	Flint glass	1.72	0.10	0.20	0.10	10.0

Note. The figures for crystal quartz refer to the ordinary ray. The figures for crown glass and flint glass refer to the particular types UBK and SF1, respectively.

a guide only because the actual figures vary greatly from one type to another. Any glass prism gives higher dispersion than quartz down to its transmission limit, which is between about 370 nm for dense (flint) glass and 330 nm for light (crown) glass. The last two columns of the table give the linear dispersion and reciprocal dispersion obtained with a camera lens of focal length 50 cm, typical of a medium-sized instrument. For a large instrument f may be 150 cm, and the reciprocal dispersion is accordingly about three times smaller. The figures for linear dispersion assume the lenses to be achromatic, so that f is the same for all wavelengths. However, it is quite common to use simple lenses, particularly in the ultra-violet, because of the expense of quartz–fluorite achromats, and to compensate for the variation in focal length with wavelength by tilting the detector plane, as shown in Fig. 5.12. A wavelength range $\Delta\lambda$ occupying Δl mm as measured perpendicular to the optic axis is then spread over a length $\Delta l/\cos\varphi$, so that the actual reciprocal dispersion is $\cos\varphi\, d\lambda/dl$. As φ is usually 60–70°, this represents an improvement by a factor of 2–3.

The angular dispersion of a quartz prism at 200 nm is about the same as that of a diffraction grating with 1200 lines mm^{-1} used in the second order, but the angular dispersion of a grating changes only slowly with wavelength within one order. It is apparent from the table that the performance of prisms falls below that of gratings very rapidly as the wavelength is increased.

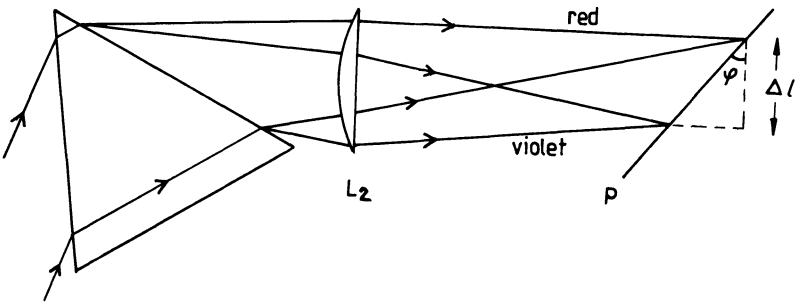

Figure 5.12 Illustration of plate tilt compensating for chromatic aberration of the simple camera lens L_2. The plate P is tilted through an angle φ from the normal to the optic axis.

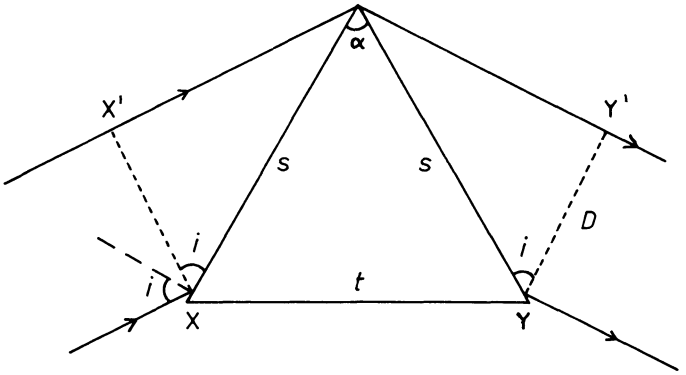

Figure 5.13 Geometry of prism at minimum deviation to show $D = t\cos i/(2\sin\alpha/2)$ and $\mathscr{R} = t\,dn/d\lambda$, where t is the prism base and s is its slant height.

5.7 Resolving power

From equation 5.10 the resolving power of a prism is $D(d\theta/d\lambda)$, where D is the diameter of the limiting aperture. Assuming this to be determined by the size of the prism, it can be seen from Fig. 5.13 that at minimum deviation $D = s\cos i$, where s is the slant height of the prism. In terms of the prism base t,

$$t = 2s\sin\left(\frac{\alpha}{2}\right) \quad \text{giving} \quad D = \frac{t\cos i}{2\sin(\alpha/2)}$$

From equations 5.14 and 5.13,

$$\frac{d\theta}{d\lambda} = \frac{2\sin(\alpha/2)}{[1-n^2\sin^2(\alpha/2)]^{1/2}}\frac{dn}{d\lambda} = \frac{2\sin(\alpha/2)}{(1-\sin^2 i)^{1/2}}\frac{dn}{d\lambda}$$

and we end up with the relatively simple expression

$$\mathscr{R}=\frac{\lambda}{\delta\lambda}=t\frac{\mathrm{d}n}{\mathrm{d}\lambda} \tag{5.16}$$

for the maximum resolving power of a prism spectrometer. This relation can be quite easily reconciled with the ideas developed in section 5.2. If XX′ in Fig. 5.13 is the incident wavefront and YY′ the refracted wavefront, then the optical path between the two wavefronts is by definition the same whether it is measured along XY, X′Y′ or any ray between. We require this optical path, which is obviously nt as measured along XY, to differ by λ for the two wavelengths λ and $\lambda+\delta\lambda$:

$$\delta(nt)=t\frac{\mathrm{d}n}{\mathrm{d}\lambda}\delta\lambda=\lambda \Rightarrow \mathscr{R}=t\frac{\mathrm{d}n}{\mathrm{d}\lambda}$$

From Table 5.1 it can be seen that the dispersion $\mathrm{d}n/\mathrm{d}\lambda$ in the visible/ultra-violet is of order $3\times10^{-4}\,\mathrm{nm}^{-1}$, or $300\,\mathrm{mm}^{-1}$. Taking 100 mm as an upper limit for the prism base, one arrives at 30 000 as a reasonable rough limit for the resolving power of a prism.

5.8 Prism materials

The various types of glass available cover the spectral region between about 2.5 μm and 350 nm. Quartz, either crystalline or as fused silica, extends the range a little at both ends, to about 3.5 μm and 170 nm. Crystalline quartz has the higher dispersion, but prisms made from it have to be specially constructed to allow for double refraction and optical activity (a composite prism of right- and left-handed quartz is required to compensate the second effect). Special – and different – varieties of fused silica are required for good transmission in the furthest ultra-violet and infra-red, respectively.

The high dispersion of a transparent material near its short wavelength cut-off is remarked on in section 5.6. The same is true near the long wavelength cut-off where $\mathrm{d}n/\mathrm{d}\lambda$ increases as λ increases. The sign of $\mathrm{d}n/\mathrm{d}\lambda$ is the same at both ends of the transmission region, so the deviation is always greater for shorter wavelengths. In the case of quartz $\mathrm{d}n/\mathrm{d}\lambda$ goes through a minimum at about 1.5 μm, and the 600 nm value is regained at 3 μm.

Prism spectrometers in the vacuum ultra-violet beyond the silica cut-off are seldom useful instruments. It is possible to make small prisms of calcite, lithium fluoride, etc., that transmit down to about 130 nm [10, 11], but their use for practical purposes is restricted to predispersing elements for echelle gratings. In the infra-red the various alkali halides are used beyond the silica transmission limit, starting with rock salt (up to 16 μm) and ending with cesium iodide at about 60 μm. They all have the disadvantage of being hygroscopic to a greater or lesser extent. References [7–9] give useful data on transmission, refractive

index and dispersion as a function of wavelength for the infra-red materials. The reasons for persisting with prisms into the infra-red are that gratings often present severe order-sorting problems, and their potentially higher resolving power may be unusable because of intensity limitations.

5.9 Image defects

5.9.1 LINE CURVATURE

With the slit at the focus of the collimator lens, all rays from the centre of the slit pass through the prism parallel to a principal section, but rays from points on the slit above and below the centre pass through at a small inclination to it. The path through the prism is effectively longer for these rays, and they are deviated more. The slit image becomes an arc with the ends curved towards shorter wavelengths. The radius of curvature is of the same order as the focal length of the camera lens, but as it depends on refractive index it increases towards shorter wavelengths. A curved image on a photographic plate may not matter, but it limits the useful slit height for a photoelectric detector with associated exit slit. Curved slits are sometimes used to increase the useful slit height.

5.9.2 LENS ABERRATIONS

In addition to the curvature introduced by the prism, the slit image is affected by all the normal lens aberrations from the collimator and camera lenses. Of these, spherical aberration is insignificant at the small *f*-numbers normally used, and chromatic aberration in the camera lens is hardly a defect in this context, since it may be compensated by tilting the observation plane as described in section 5.6. However, chromatic aberration in the collimator lens means that the wavefront incident on the prism cannot be plane for all wavelengths, as a result of which the prism introduces coma. The relative aperture of the collimator should be limited to about *f*/10 to keep this within tolerable limits. The camera lens has to work at a rather wide angle in the direction of the dispersion if a reasonably wide section of the spectrum is to be imaged, so the off-axis aberrations of astigmatism and coma are important for this lens. Both defects need to be kept within the film grain/slit width tolerance by limiting the *f*-number; coma is objectionable because of its asymmetry, and astigmatism is an undesirable property for any instrument that may be required for cross-dispersion or spatial resolution. The final aberration is field curvature. Photographic instruments used to have curved plateholders to match this curvature, but many modern instruments achieve a flat field suitable for photoelectric scanning by 'bending' the lens, which means adjusting the curvature of the two faces while maintaining the same focal length.

These remarks apply to the normal run of *f*/10 to *f*/25 instruments but not, of course, to the very high aperture Raman spectrometers, for which lens designs of great sophistication are required.

5.10 Merits of prism spectrometers

The principal advantages of prism instruments are that they are relatively cheap and simple to construct and adjust. They are stigmatic, by contrast with concave (though not plane) gratings, and are therefore suitable auxiliary instruments for interferometers – e.g. for cross-dispersion with Fabry–Perot interferometers. In the quartz ultra-violet region, where they are at their most efficient, their light losses are likely to be less than those of grating instruments. Finally, the spectrum is quite unambiguous, with no problems of overlapping orders.

The most important disadvantage is undoubtedly the rather low dispersion and resolving power attainable. Prism instruments share with gratings a lower throughput than can be obtained with an interferometer for a given resolving power. They cannot be used in the far infra-red or far ultra-violet regions of the spectrum. In some applications the very non-linear wavelength scale is a disadvantage.

To summarize, prism instruments are cheap answers to undemanding requirements and have useful applications as cross-dispersers and survey instruments.

References

GENERAL

1. Born, M. and Wolf, E. (1965) *Principles of Optics*, 2nd edn, Pergamon Press, Oxford.
2. Bousquet, P. (1971) *Spectroscopy and its Instrumentation*, Hilger, London.
3. Ditchburn, R.W. (1963) *Light*, Blackie, London and (3rd edn, 1976) Academic Press, London.
4. Harrison, G.R., Lord, R.C. and Loofbourow, J.R. (1948) *Practical Spectroscopy*, Prentice-Hall, Englewood Cliffs, New Jersy.
5. Jenkins, F.A. and White, H.E. (1950) *Fundamentals of Physical Optics*, McGraw-Hill, New York.
6. Sawyer, R.A. (1961) *Experimental Spectroscopy*, Dover, New York.

THE INFRA-RED REGION

7. Chantry, G.W. (1984) *Long-wave Optics* (2 Vols), Academic Press, New York.
8. Conn, G.K.T. and Avery, D.G. (1960) *Infrared Methods*, Academic Press, New York.
9. Smith, R.A., Jones, F.E. and Chasmar, R.P. (1968) *Detection and Measurement of Infra-red Radiation*, Oxford University Press.

THE ULTRAVIOLET REGION

10. Samson, J.A.R. (1967) *Techniques of Vacuum Ultra-violet Spectroscopy*, Wiley, New York.
11. Zaidel, A.N. and Schreider, E.Y.A. (1970) *Vacuum Ultraviolet Spectroscopy*, Humphrey Science, Ann Arbor, Michigan.

SIX

Diffraction gratings

A diffraction grating consists of a large number of close, equally spaced 'grooves' on a plane or concave surface. These may actually be ruled on the surface, or formed by replicating a ruled surface, or generated by etching a suitably prepared coating as in holographic gratings. Collimated light from a slit set parallel to the grooves is transmitted by or reflected from the periodic structure formed by the grooves, which behave as a large number of coherent sources. For any given wavelength the secondary wavelets from them interfere constructively at certain angles and destructively at others. Maxima of intensity appear in the focal plane of a camera lens at angles corresponding to the interference maxima. Concave gratings differ from plane gratings in that the grooves are ruled on a concave mirror blank, and the grating acts as its own collimator and camera lens.

6.1 Plane grating: condition for maxima

Figures 6.1 and 6.2 show a plane wave incident at an angle θ on a transmission and a reflection grating, respectively, each with groove spacing d. The path difference between contributions from adjacent grooves to a wave diffracted at angle θ' is $l = d(\sin\theta + \sin\theta')$, where θ' is taken to be positive if it is on the same side of the normal as θ and negative if it is on the opposite side. For light of wavelength λ the condition that contributions from all rulings interfere constructively is

$$m\lambda = l = d(\sin\theta + \sin\theta') \tag{6.1}$$

where m is an integer, the order number. Zero order, $\theta = -\theta'$, corresponds to straight-through transmission or specular reflection. If θ' is negative and $|\theta'| > \theta$, the order number is negative, as shown in Fig. 6.2(c).

In practice transmission gratings are scarcely ever used outside teaching laboratories. In this chapter gratings are to be understood as reflection gratings, although the same theory is applicable to both.

Evidently a given wavelength λ can have several maxima at different values of θ' corresponding to different orders m. The complete spectrum from a given

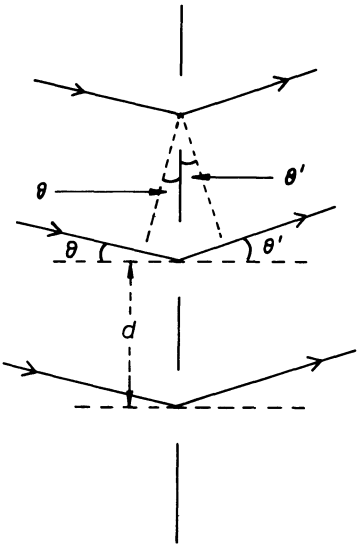

Figure 6.1 Angles of incidence (θ) and diffraction (θ') for a plane wave and a transmission grating of spacing d.

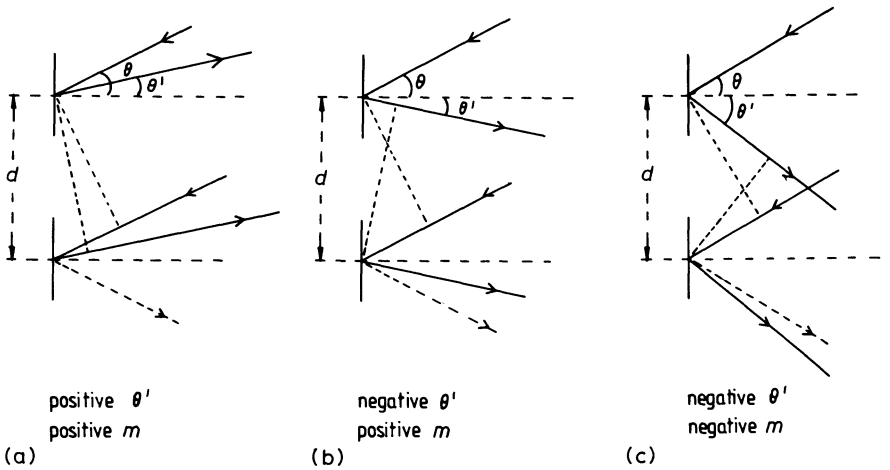

Figure 6.2 Angles of incidence (θ) and diffraction (θ') for a reflection grating, showing positive and negative orders. The broken arrow in each case is the direction of specular reflection ($-\theta' = \theta$). In (a) the diffracted ray is the same side of the normal as the incident ray, so θ' and m are both positive. In (b) and (c) they are on opposite sides, so θ' is negative and m is positive for $|\theta'| < \theta$ and negative for $|\theta'| > \theta$.

source is repeated, possibly several times, and the long-wavelength end of one order may well overlap the short-wavelength end of another – for example, the second order of 300 nm falls in the same position as the first order of 600 nm. Short wavelengths are deviated least – i.e. they are closest to zero order; this is the opposite way round to a prism.

Equation 6.1 gives the necessary information for measurement of wavelengths, and the angular dispersion can be obtained by differentiating it. From section 5.2 we anticipate a diffraction-limited instrument function and a resolving power of Nm, where N is the number of rulings. In fact we already know most of the important characteristics of gratings. All the same, it is worth deriving the intensity distribution from first principles in order to link up the definitions and to see the influence of other factors that affect performance, such as the size and shape of the grooves.

6.2 Intensity distribution from ideal diffraction grating

Figure 6.3 shows a plane monochromatic wave incident at an angle θ on an arbitrary periodic structure, and as we have already seen the path difference l for corresponding points P on adjacent rulings is

$$l = d(\sin\theta + \sin\theta')$$

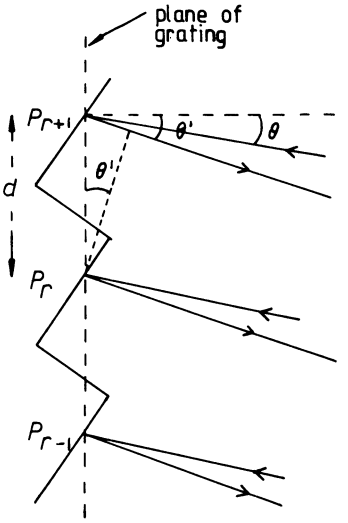

Figure 6.3 Diffraction of plane wave for arbitrary periodic structure of spatial period d. P_r are corresponding points in successive periods.

The corresponding phase difference is

$$\beta = \frac{2\pi l}{\lambda} = \frac{2\pi d}{\lambda}(\sin\theta + \sin\theta') \tag{6.2}$$

We take the zero of phase to be at P_1 at one end of the grating. If the amplitude of an element at each point P is δa, the complex amplitude of the contribution to the diffracted wave from the element P_r is

$$\delta a\, e^{-ir\beta}$$

The complex amplitude A for the whole grating area is obtained by integrating this expression over one spatial period and summing over all rulings:

$$A = \sum_{r=1}^{N} e^{-ir\beta} \int_{\text{groove}} da$$

The integral is independent of r, so it can be taken outside the sum. It is a function of θ' and λ, but we postpone consideration of it for the moment, and simply write it as a complex amplitude a:

$$a = \int_{\text{groove}} da.$$

Therefore,

$$A = a \sum_{r=1}^{N} e^{-ir\beta} = a\frac{1 - e^{-iN\beta}}{1 - e^{-i\beta}}$$

The intensity is obtained by multiplying A by its complex conjugate

$$A^* = a^*\frac{1 - e^{iN\beta}}{1 - e^{i\beta}}$$

giving

$$I = AA^* = i_0\frac{1 - \cos N\beta}{1 - \cos\beta} = i_0\frac{\sin^2(N\beta/2)}{\sin^2(\beta/2)} = i_0\frac{\sin^2 N\delta}{\sin^2\delta} \tag{6.3}$$

where

$$i_0 = aa^* \quad \text{and} \quad \delta = \frac{\beta}{2} = \frac{\pi d}{\lambda}(\sin\theta + \sin\theta') \tag{6.4}$$

i_0 is the intensity envelope from a single groove and can usually be taken to vary relatively slowly with θ', as will be shown below. The term $\sin^2(N\delta)/\sin^2\delta$ has principal maxima whenever $\delta = m\pi$, where m is a positive or negative integer:

$$\left(\frac{\sin N\delta}{\sin\delta}\right)^2 \Rightarrow \left(\frac{N\delta}{\delta}\right)^2 = N^2 \quad \text{for} \quad \delta = m\pi$$

The condition for the principal maxima is therefore the one anticipated in

equation 6.1:

$$d(\sin\theta+\sin\theta')=m\pi$$

Between these principal maxima are secondary maxima and minima corresponding to the N-times more rapid oscillations of $\sin N\delta$, as shown schematically in Fig. 6.4. The minima occur for $\delta=(p/N)\pi$, where p is another integer, except that whenever p/N is itself an integer we get back to a principal maximum. There are $N-1$ of these minima between each pair of principal maxima at angles θ' defined by

$$d(\sin\theta+\sin\theta')=(p/N)\lambda$$

If N is large, so that $\sin\delta$ changes slowly compared with $\sin N\delta$, the $N-2$ secondary maxima fall almost half-way between the minima. Their intensity is $i_0/\sin^2\delta$, so as $\sin\delta$ increases towards its maximum value of unity half-way between principal maxima the intensities of the secondary maxima fall towards i_0, which is a fraction $1/N^2$ of the principal maxima. N is of order 10^4–10^5 for an ordinary grating, and even for an echelle it is greater than 1000, so it is safe to ignore all secondary maxima except for those close to the principal maxima.

These principal maxima are the 'spectral lines'. To show that they have the anticipated diffraction line shape, we have to compare equation 6.3 with equation 5.9:

$$I=I_0\left(\frac{\sin\alpha}{\alpha}\right)^2 \quad \text{where} \quad \alpha=\frac{\pi D\varphi}{\lambda}$$

and at first sight they do not look very much alike. But φ in equation 5.9 is the

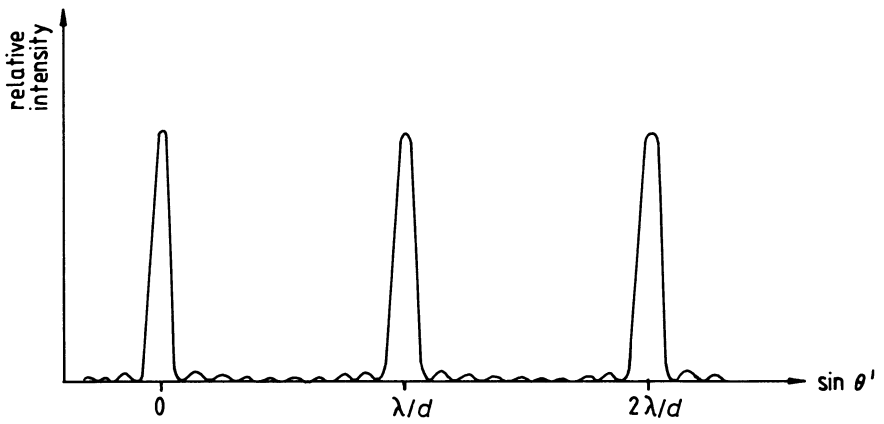

Figure 6.4 Schematic representation of grating intensity distribution. The figure is drawn for $\theta=0$, giving principal maxima at $\sin\theta'=m\lambda/d$, and for nine rulings, giving eight subsidiary maxima between them. In practice $N\sim10^4$–10^5, and only the subsidiary maxima adjacent to the principal maxima are significant.

angular spread around a principal maximum, so $\varphi = \theta' - \theta_0$, where θ_0 satisfies equation 6.1,

$$d(\sin\theta + \sin\theta_0) = m\lambda$$

φ is necessarily small, and δ in equation 6.4 can be expanded as

$$\delta = \frac{\pi d}{\lambda}(\sin\theta + \sin\theta_0 \cos\varphi + \cos\theta_0 \sin\varphi) \simeq \frac{\pi d}{\lambda}(\sin\theta + \sin\theta_0 + \varphi\cos\theta_0)$$

Therefore,

$$\delta = m\pi + \frac{\pi d \cos\theta_0}{\lambda}\varphi \quad \text{and} \quad N\delta = Nm\pi + \frac{\pi N d \cos\theta_0}{\lambda}\varphi$$

Since $Nd\cos\theta_0 = W\cos\theta_0 = D$, as shown in Fig. 6.5, we have

$$\sin N\delta = \sin(\pi D\varphi/\lambda) \equiv \sin\alpha$$

and

$$\sin\delta = \sin(\pi D\varphi/N\lambda) \equiv \sin\left(\frac{\alpha}{N}\right) \simeq \frac{\alpha}{N}$$

Equation 6.3 becomes

$$I = i_0 N^2 \left(\frac{\sin\alpha}{\alpha}\right)^2$$

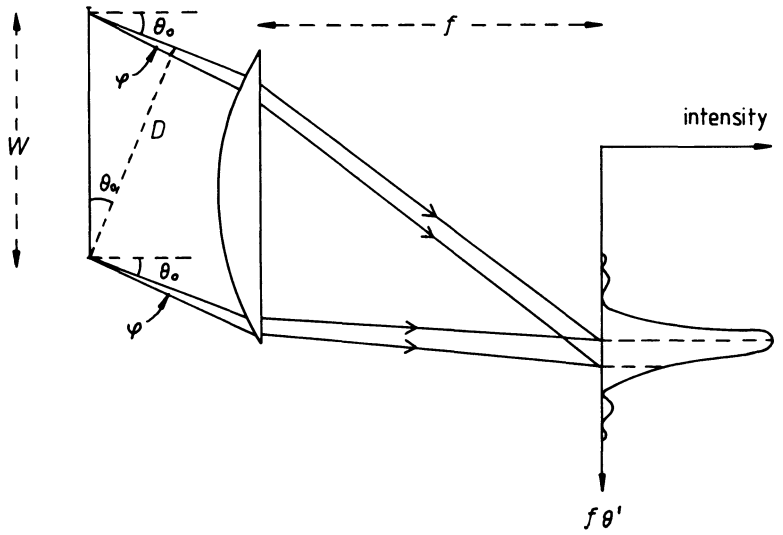

Figure 6.5 Intensity distribution around a principal maximum for light incident normally on a grating of width W. θ_0 satisfies the condition $m\lambda = d\sin\theta_0$, and φ is the angular spread around θ_0. From the figure $D = W\cos\theta_0$.

The only difference between this and equation 5.9 is the factor N^2, which has crept in only because in Chapter 5 we took unit amplitude to represent the flux incident on the whole aperture, whereas in this section unit amplitude referred to one groove.

There remains to be considered the form of the integral $a = \int_{\text{groove}} \mathrm{d}a$ giving the intensity envelope. In the bar-and-space model shown in Fig. 6.6, each groove is a 'top-hat' function, reflecting or transmitting uniformly over its width b. i_0 is then the Fraunhofer diffraction distribution for a slit of width b:

$$i_0 = \text{const.}\left(\frac{\sin\beta}{\beta}\right)^2 \quad \text{where} \quad \beta = \frac{\pi b \sin\theta'}{\lambda} \tag{6.5}$$

This distribution has its first minimum at $\sin\theta' = \pm\lambda/b$ for any given wavelength. Since $b \simeq d/2$, the central maximum of the intensity envelope has an angular spread only a little larger than the distance between two principal maxima. However, this is a very rough approximation to a real grating. Modern gratings are nearly always 'blazed' to direct a large proportion of the flux at a particular diffraction angle, as described in section 6.6. There is no simple way of evaluating the groove integral for a real grating, but the intensity envelope it produces is always a relatively slowly varying function of angle.

6.3 Dispersion and resolving power

The angular dispersion is found by differentiating equation 6.1 to be

$$\frac{\mathrm{d}\theta'}{\mathrm{d}\lambda} = \frac{m}{d\cos\theta'} \tag{6.6}$$

Near the grating normal, where $\cos\theta' \simeq 1$, the dispersion is m/d and is nearly constant, giving an almost linear wavelength scale. Gratings are very rarely used in this configuraion (section 6.5), but with $\theta' \simeq 30°$ the departure from linearity is not very great. Standard ruled gratings for the visible and ultra-violet usually have 600 or 1200 lines mm^{-1}. The first of these, used in the first order, has an

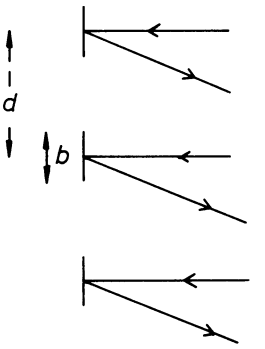

Figure 6.6 'Bar-and-space' model of a grating. The bars have width b and spacing d.

angular dispersion of 6×10^{-4} rad nm^{-1}, which gives a reciprocal dispersion of 1.6 nm, or 16 Å, per mm with a 1 m spectrometer. A 1200 line mm^{-1} grating in the second order with the same spectrometer gives 0.4 nm mm^{-1} which, as remarked in section 5.6, is comparable with a prism instrument of the same focal length at a wavelength just above its cut-off. The same grating and the same order in a 3 m instrument gives 0.13 nm mm^{-1}.

For the resolving power we have the three equivalent expressions of section 5.2 (equations 5.8 and 5.10):

$$\mathscr{R} = \frac{L}{\lambda} = Nm = D\frac{\mathrm{d}\theta'}{\mathrm{d}\lambda} \tag{6.7}$$

(The second of these can be obtained directly from the third, using $\mathrm{d}\theta'/\mathrm{d}\lambda$ from equation 6.6 and $D = Nd\cos\theta'$.) It is also useful to express $\mathscr{R}$ in terms of the grating width, W,

$$\mathscr{R} = \frac{L}{\lambda} = \frac{W(\sin\theta + \sin\theta')}{\lambda} = W\frac{m}{d} \tag{6.8}$$

showing that for a given wavelength the resolving power is determined by W and the ratio m/d; this last can be the same for a fine ruling in a low order as for a coarse ruling in a high order, an important principle which is put into practice in the echelle grating. The importance of the grating width is brought out clearly in the commonly used Littrow-type mountings (section 6.5 and Fig. 6.11), for which

$$\theta \simeq \theta' \Rightarrow \mathscr{R} = \frac{2W\sin\theta}{\lambda} \tag{6.9}$$

For a typical value of θ, 30°,

$$\mathscr{R} = W/\lambda$$

and the ultimate limit, as the grating is turned towards grazing incidence, is $2W/\lambda$. Similarly, the resolution limit is given by

$$\delta\sigma \simeq 1/W \tag{6.10}$$

where W must be in centimetres to obtain σ in wavenumbers.

There are practical limits to the width over which a grating can be ruled with sufficient accuracy, and it sometimes pays actually to decrease the effective width by blanking off bad parts of the surface. A 1200 line mm^{-1} grating, 200 mm wide, used in second order has a theoretical resolving power of almost half a million. Such large resolving powers are seldom realized in practice for various reasons, some of which are mentioned in section 6.6, but resolving powers of a few hundred thousand are quite feasible from the infra-red to the vacuum ultra-violet region, and echelle gratings (section 6.7) can approach a million.

So far it has been assumed that resolving power is diffraction-limited. As stated in section 5.3, the actual instrument function is the convolution of the

diffraction distributon with the geometric width of the slit image. Assuming for simplicity that the spectrometer has unit magnification, so that the slit and its image have equal widths, the optimum slit width is given by equation 5.11 as $w_0 = f\lambda/D$. The resolution is degraded if the slit is appreciably wider than this. For a grating in a Littrow-type mounting, $D = W\cos\theta$, and the relation between W and the limiting resolving power is given by equation 6.9 as

$$\frac{\lambda}{W} = \frac{2\sin\theta}{\mathscr{R}}$$

Therefore,

$$\frac{w_0}{f} = \frac{\lambda}{D} = \frac{\lambda}{W\cos\theta} = \frac{2\tan\theta}{\mathscr{R}} \tag{6.11}$$

Evidently w/f has to be small if the resolution of a large grating is to be fully exploited. Mechanical constraints on the slit set a lower limit of about 10 μm on w, and about the same limit is imposed by plate grain in photographic use, so equation 6.11 can be taken to specify a minimum focal length

$$f \geqslant \frac{\mathscr{R}}{2\tan\theta} \times 10^{-5} \simeq 10^{-5}\mathscr{R} \tag{6.12}$$

with f in m. This relation explains why high-resolution spectrometers are so large. A 1 m instrument will go only up to $\mathscr{R} \simeq 100\,000$, and the next standard length, 3 m, is already, with its necessary mountings and adjustments, a large and heavy instrument. The largest spectrographs, with $\mathscr{R}$ around half a million, have focal lengths of 5–10 m.

The throughput T of a spectrometer, which determines the power reaching the detector, is proportional to $wh(D/f)^2$ (equation 5.12). Thus, a long focal length carries a flux penalty in addition to its inconvenience. In many actual experiments, particularly in the infra-red and far ultra-violet where source radiance is likely to be low, the throughput has to be increased in the only way possible for a given instrument – by opening the slit up beyond the optimum value, at the expense of resolving power. The slit-limited resolution $\delta\lambda$ is given by

$$\delta\lambda = \frac{\mathrm{d}\lambda}{\mathrm{d}l}\delta l = \frac{1}{f}\frac{\mathrm{d}\lambda}{\mathrm{d}\theta}w = \frac{d\cos\theta' w}{m f} \quad \text{from equation 6.6}$$

Again for the Littrow-type mounting, $\theta = \theta'$ and $m\lambda = 2d\sin\theta$, giving

$$\mathscr{R}_{\text{slit}} = \frac{\lambda}{\delta\lambda} = \lambda\frac{m}{d\cos\theta'}\frac{f}{w} = \frac{2f\tan\theta}{w} = \frac{w_0}{w}\mathscr{R} \tag{6.13}$$

Not surprisingly, the resolving power is degraded by the factor w_0/w from the grating or diffraction-limited value of equation 6.11.

Equation 6.13 brings out the inverse relation between resolving power and throughput. This is often quantified by taking their product $\mathscr{R}T$ to define a figure of merit which can be used to compare different types of spectrometer. For

an entrance slit of dimensions $w \times h$, T is given by equation 5.12:

$$T = wh\frac{\pi}{4}\left(\frac{D}{f}\right)^2$$

Replacing the circular area $\pi D^2/4$ by the actual projected grating area, of width $W\cos\theta$ and height H, we have

$$T = wh\,WH\cos\theta/f^2$$

Therefore,

$$\mathscr{R}_{\text{slit}}T = 2\tan\theta\frac{h}{f}WH\cos\theta = 2\beta\,WH\sin\theta \simeq \beta WH \tag{6.14}$$

where β is the angular height of the slit as seen from the grating. In practice β is limited by slit-image curvature and alignment problems to about 1/100.

An interesting sideline on figure of merit is obtained by using the Littrow version of equation 6.1, $2\sin\theta = m\lambda/d$, to rewrite equation 6.14 as

$$\mathscr{R}_{\text{slit}}T = \beta\,WH\frac{m\lambda}{d} = \beta\, m\lambda\, NH$$

$\mathscr{R}T$ is thus proportional to the total length NH of the grooves, or the distance travelled by the ruling diamond.

6.4 Concave gratings

It was first shown by Rowland in 1882 that if a grating is ruled on a concave mirror instead of on a plane surface it can act as its own collimator and camera lens. The spectrometer is reduced to the three essential elements of slit, grating and detector, which is particularly advantageous in the far ultra-violet where lenses do not exist and mirror reflectivities are very low. Rowland showed that if the slit and grating both lie on a circle whose diameter is equal to the radius of curvature of the grating, and if the grating is tangent to this circle, then for small apertures (the equivalent of paraxial rays in a lens system) the spectral lines are in focus on the circle. The positions of the lines are given, as for the plane grating, by

$$d(\sin\theta + \sin\theta') = m\lambda \tag{6.1}$$

where d is the distance between rulings as measured along a chord across the grating arc.

The Rowland condition can be derived quite easily for rays in a plane perpendicular to the grating if the width of the grating is small enough for it to be considered coincident with the circle instead of tangent to it. Figure 6.7 shows two adjacent rulings A and B of a grating whose centre of curvature is at C. Rays from the slit S are incident at angle θ, are diffracted at angle θ', and intersect at P.

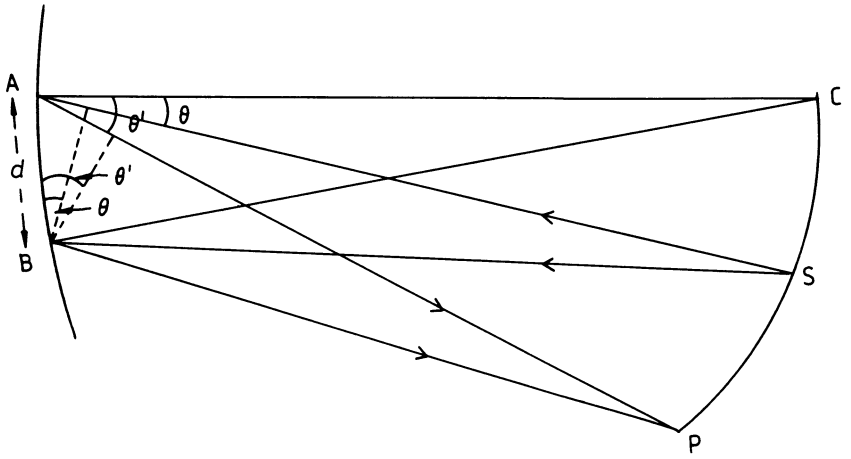

Figure 6.7 Angles of incidence θ and diffraction θ' for concave grating. C is the centre of curvature of the grating, A and B are adjacent rulings (with AB much exaggerated), S is the slit and P is its image.

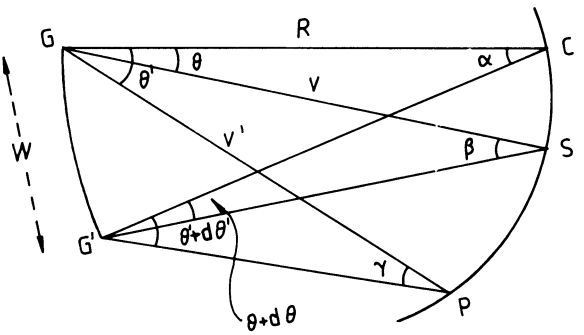

Figure 6.8 Ray diagram for concave grating. The grating GG′ of width W and radius of curvature R subtends angles α, β, γ at C, S, P, respectively. v and v' are the object (slit) and image distances, respectively.

To a first approximation the path difference between the two incident rays SA and SB is $d \sin\theta$, and that between the two diffracted rays AP and BP is $d \sin\theta'$. Equation 6.1 therefore defines the condition for a principal maximum at P so far as this little bit of the grating is concerned. We now need to show that all rays from S that satisfy equation 6.1 intersect at the same point P, wherever they hit the grating. Figure 6.8 shows the grating GG′ of width W and radius of curvature R, with S at a distance v and P at distance v'. The angles of incidence and diffraction are θ and θ' at G. If the small aperture condition $W \ll v, v'$ is fulfilled, the corresponding angles at G′ can be written as $\theta + d\theta$ and $\theta' + d\theta'$

where equation 6.1 requires

$$\cos\theta\,\mathrm{d}\theta+\cos\theta'\,\mathrm{d}\theta'=0$$

The angles α, β and γ subtended by GG′ at C, S and P, respectively, are given by

$$\alpha=\frac{W}{R}\qquad \beta=\frac{W\cos\theta}{v}\qquad \gamma=\frac{W\cos\theta'}{v'}$$

It can be seen from the geometry of the figure that

$$\theta+\alpha=\theta+\mathrm{d}\theta+\beta \quad\text{and}\quad \theta'+\alpha=\theta'+\mathrm{d}\theta'+\gamma$$

Therefore,

$$\mathrm{d}\theta=\alpha-\beta \quad\text{and}\quad \mathrm{d}\theta'=\alpha-\gamma$$

The condition that rays from S to P satisfy equation 6.1 over the whole width of the grating is therefore

$$\cos\theta\left(\frac{1}{R}-\frac{\cos\theta}{v}\right)+\cos\theta'\left(\frac{1}{R}-\frac{\cos\theta'}{v'}\right)=0 \tag{6.15}$$

In mountings based on the Rowland circle this is satisfied by making both expressions in parentheses identically zero – i.e. $v=R\cos\theta$ and $v'=R\cos\theta'$. As shown in Fig. 6.9, S and P then lie on a circle of diameter R, which is the Rowland circle. This is the simplest and most aberration-free way of mounting the concave grating, but equation 6.15 can be satisfied by other geometric arrangements. Actual mountings are discussed in section 6.5.

Since plane and concave gratings obey the same basic grating equation, the expressions already derived for dispersion and resolving power still hold. The

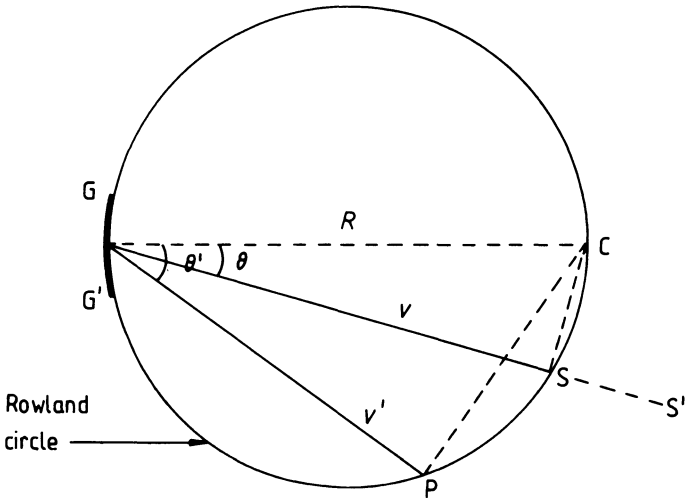

Figure 6.9 Geometry of Rowland circle. The letters have the same meanings as in Figs 6.7 and 6.8. S′ is the Sirks focus (see text).

only modification is that the focal lengths of the lenses are replaced by the radius of curvature of the grating, so that the reciprocal dispersion becomes

$$\frac{d\lambda}{dl} = \frac{1}{R}\frac{d\lambda}{d\theta'} = \frac{d\cos\theta'}{R} \tag{6.16}$$

The 'standard' values for R are 1, 3 and 6 or 6.6 m, the last two being very large instruments. In fact, gratings of up to 10 m radius have been used, but the angular aperture, and hence the light throughput, of these large instruments is very small because ruling a good large grating is even harder if the grating is concave than if it is plane.

Astigmatism is an important defect of concave gratings that is not shared by plane gratings. The imaging properties of a concave grating are those of a concave mirror used off-axis. If the slit is accurately parallel to the grating rulings, each point of the slit is imaged at the optimum horizontal focus as a short vertical line, and there need be no loss of definition. In general, however, the flux density is reduced, and the point-to-point correspondence between the slit and its image is lost. This means that it is not possible to resolve spatially an image of the source, a diaphragm or a set of interference fringes on the slit, although it is sometimes possible to work with the source image or diaphragm at the Sirks focus; this is the position S′ on Fig. 6.9 for which a point object is imaged on the Rowland circle as a sharp *horizontal* line. It is actually possible to eliminate astigmatism altogether by using an auxiliary concave mirror, as is done in the Wadsworth mounting described in the next section.

The measure of astigmatism is the length z of the line image from a point on the slit. z is proportional to the vertical aperture H/R, where H is the length of the rulings, and it has a minimum value of $2H\sin^2\theta$ for $\theta' = \theta$; this works out to be about $H/2$ for the second order of a 1200 line mm^{-1} grating in the visible region. The flux density in the central part of the image would be little reduced if it were possible to work with a slit length approximately equal to z, but this is usually too long to be practicable. Even if a slit some centimetres long could be illuminated, the useful length is limited by the curvature of the image, an aberration which arises from the different angles of incidence of the off-axis rays from the ends of the slit and which has already been described as the limiting factor for slit height in a prism instrument. The combination of curvature with astigmatism is unacceptable because it results in broadening of the spectral lines. Figure 6.10 shows the astigmatism as a function θ and θ'; each curve is labelled by its value of z/H. The combinations of angles that satisfy equation 6.1 for various values of $m\lambda/d$ lie on the straight lines.

6.5 Grating mountings

Many different types of mounting have been designed for diffraction gratings, some for special problems. This section describes the types in most general use;

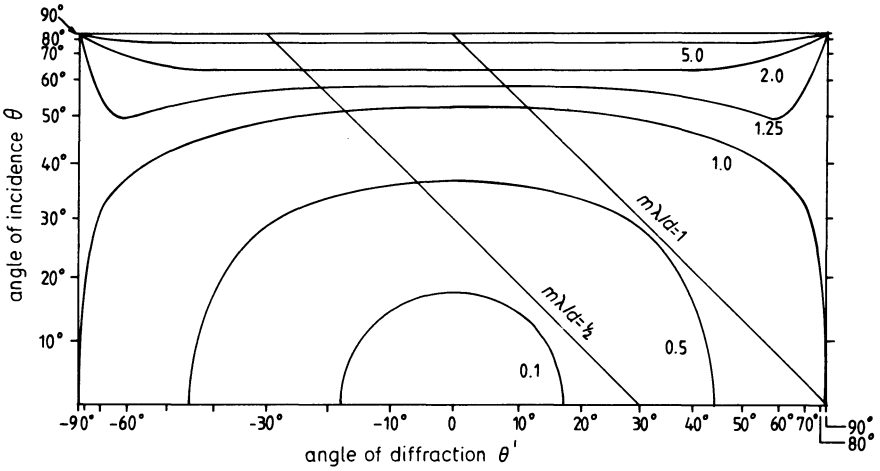

Figure 6.10 Astigmatism of concave grating (after [5]). The number by each curve gives the image length z of a point source on the slit in units of the grating ruling length H. The scales along the axes are linear in $\sin\theta$ and $\sin\theta'$, respectively. The combinations of angles that satisfy $m\lambda/d = \sin\theta + \sin\theta'$ lie on straight lines of unit slope, of which the two shown correspond to $\lambda = 833$ nm and $\lambda = 417$ nm, respectively, for the first order of a grating with 1200 lines mm^{-1}.

further details are given in [3–5]. The distinction between photographic and photoelectric detection is not as important as might at first appear, since the focal plane P of the camera lens defines either the location of the plate or the region over which the detector plus exit slit must be scanned. The alternative of scanning by rotating the grating and keeping the detector fixed is simply not sufficiently accurate for high resolution, although it is satisfactory for a monochromator.

The two most common mountings for plane gratings are both of the Littrow, or autocollimating, type, for which $\theta = \theta'$. In the true Littrow (Fig. 6.11) the half-prism of Fig. 5.9 is simply replaced by a grating. To change the spectral region it is necessary to rotate the grating, refocus the lens for the centre of the plane P, and adjust the tilt of P. The second common mounting is the Ebert type shown in Fig. 6.12. The concave mirror M both collimates and focuses the light, and since there are no lenses the mounting is achromatic. The spectral region is changed simply by rotating the grating. The arrangement shown in the figure is the out-of-plane, or up-and-over, mounting, which has the advantage of a low level of scattered light. Higher resolution, especially for photoelectric detection, is achieved by the Fastie, or in-plane, arrangement with the entrance and exit slits side by side; the effects of spectral line curvature can then be overcome by using curved slits. The variant of the Ebert-type mounting in which the single mirror is replaced by two adjacent mirrors is known as the Czerny–Turner

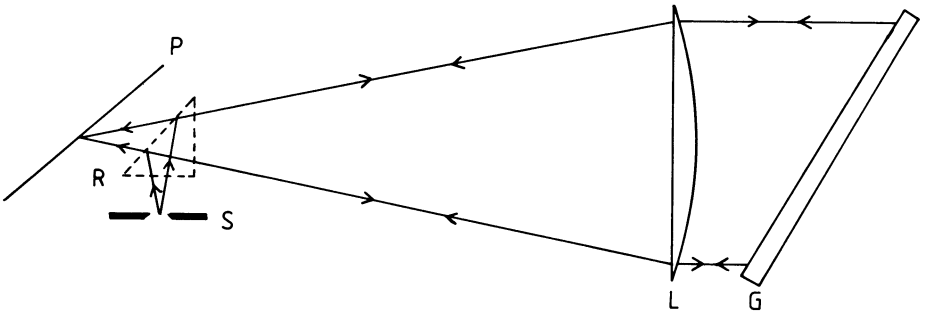

Figure 6.11 Littrow mounting of plane grating. The slit S and reflecting prism R are both above (or below) the plane of the rest of the diagram.

mounting. All Ebert-type mountings show very little coma and almost negligible astigmatism.

All common mountings of the concave grating except for the Wadsworth and the Seya use the Rowland circle in some form. Rowland's original mounting has the simplest geometry, but is now really only of historical interest. The closest modern equivalent is the Paschen–Runge mounting shown in Fig. 6.13: the slit and grating are both fixed on the Rowland circle, and the spectrum is observed round the whole arc P. Alternative fixed slits may be provided at different angles of incidence. This mounting takes up a great deal of room and is not suitable for vacuum work, but it is used in a few large research laboratories.

The most compact mounting for the concave grating is that devised by Eagle. This is a modified Littrow arrangement, using the grating in autocollimation. The axis of the spectrometer forms a chord of the Rowland circle with the grating and focal plane at its ends, as shown in Fig. 6.14. In instruments intended for the visible and near infra-red the slit is usually mounted 'off-plane', slightly above or below the axis GP and offset to the side, with a small quartz prism to reflect in the incident light. The alternative 'in-plane' mounting has the slit displaced laterally instead of vertically; this is the form usually adopted for monochromators and far ultra-violet instruments. In both cases the slit is fixed, and the spectral region is changed by rotating the grating and at the same time moving it along the spectrometer axis to maintain the correct geometry. A related rotation of the plate holder or scanning mechanism P is required. Various mechanical links for the necessary adjustment have been devised. Again, $\theta = \theta'$, and equating the angles has the additional advantage of minimizing the astigmatism. In one form or another this is the most widely used type of mounting.

The Wadsworth mounting (Fig. 6.15) gets away from the Rowland circle by using an auxiliary mirror as a collimator. The principal advantage is that the spectrum near the normal to the grating is stigmatic, and coma can be kept small provided that the slit is close to the grating. However, the distance GP is $R/2$

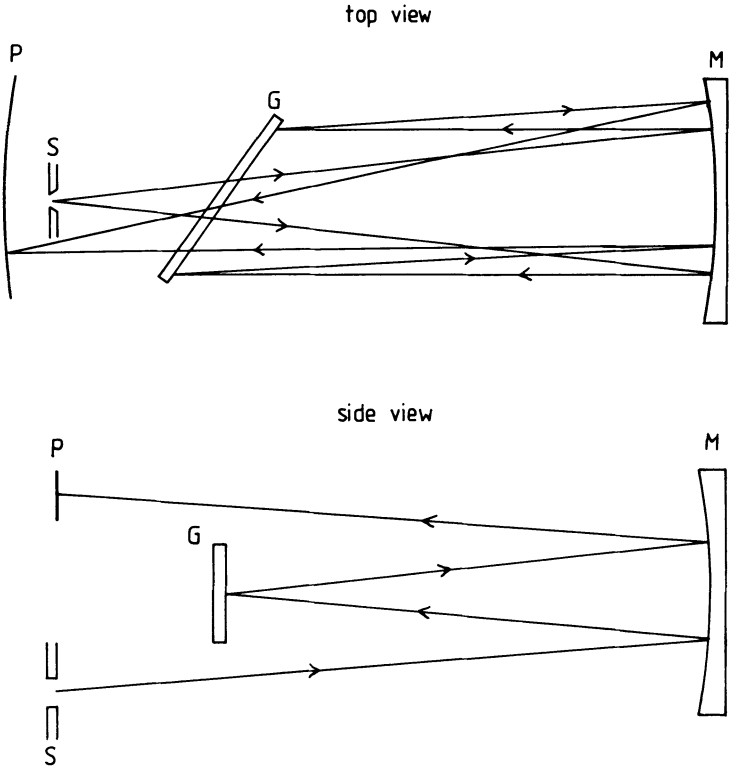

Figure 6.12 Ebert mounting of plane grating. This diagram shows the 'up-and-over' mounting with the slit S vertically below the focal plane P.

instead of R, thus halving the linear dispersion, as can be seen by putting $v = \infty$ and $\cos\theta' \simeq 1$ in equation 6.15:

$$\frac{\cos\theta}{R} + \frac{1}{R} - \frac{1}{v'} = 0 \Rightarrow v' = \frac{R}{1+\cos\theta} \simeq \frac{R}{2}$$

An additional disadvantage is that the focal plane is parabolic instead of spherical, which makes for difficulties if a long section of spectrum is to be photographed or scanned.

The priority for a monochromator is usually ease of scanning rather than high resolution, and the complicated linkages required in the concave grating mountings so far described are undesirable. A number of mountings achieve simplicity at the expense of some loss of definition, and of these probably the best known is the Seya. The exit and entrance slits of this instrument are both fixed, and the spectrum is scanned by rotating the grating about an axis through its surface. It can be shown that the optimum angle between the exit and entrance beams is

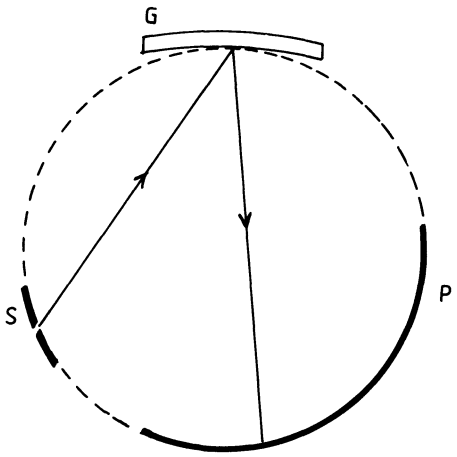

Figure 6.13 Paschen–Runge mounting of concave grating. The Rowland circle is shown dashed. The slit S, grating G and plateholder P are all fixed.

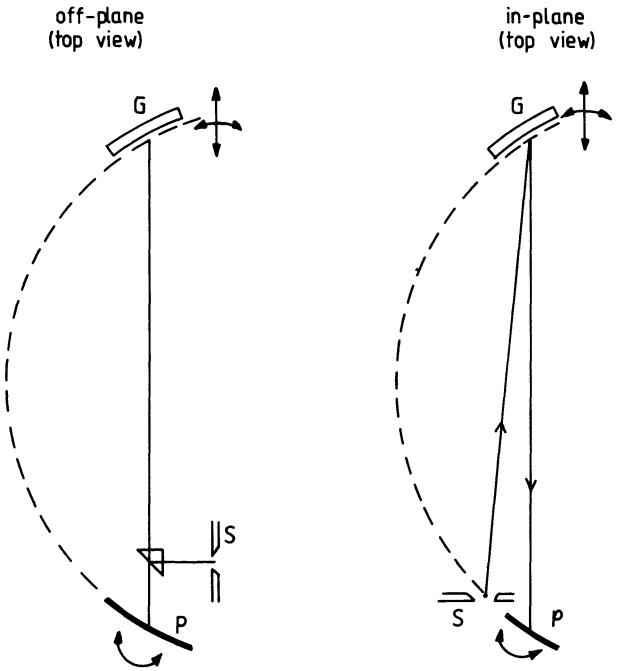

Figure 6.14 Eagle mounting of concave grating, showing both the off-plane and in-plane arrangements, with the slit S respectively above and adjacent to the focal plane. The arrows show the motions of the grating G and the plateholder P for scanning the spectrum. The broken arc is part of the Rowland circle.

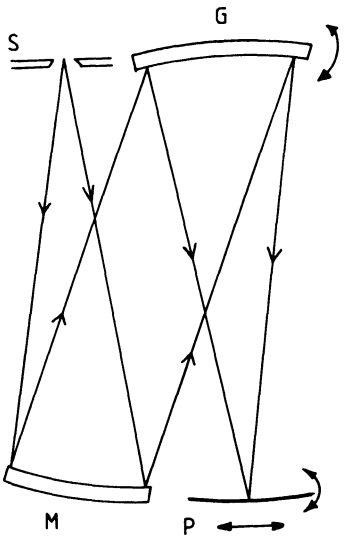

Figure 6.15 Wadsworth mounting of concave grating, using an auxiliary mirror M. The arrows again show the motions of grating G and plateholder P when scanning.

70°, and with a 1 m grating of 1200 lines mm^{-1} the focus is within 0.1 mm of the exit slit over the whole visible and near ultra-violet regions.

The angular ranges covered by these different mountings are shown in Fig. 6.16. Comparison of this with Fig. 6.10 shows how the amount of astigmatism depends on the type of mounting. The grazing incidence mounting in the top left-hand corner is used in the extreme ultra-violet and is described in section 6.7.

6.6 Production and characteristics of gratings

The early gratings were ruled on speculum metal with ruling engines based on the original Rowland engine of the 1880s. The gratings produced in the early part of the present century by such experts in the required combination of art and science as Rowland, Michelson, Wood and Strong were not greatly improved on for some 40 years, except that the speculum metal was replaced in the 1930s by a glass blank coated with a thin evaporated layer of aluminium. Two important advances of the period 1950–1970 are both developments of original ideas of Michelson's: interferometric control of ruling engines and 'holographic' gratings formed from an optical interference pattern. Almost equally important in practice was the development of grating replication techniques [7, 8].

For ruled gratings the two criteria to be met by the grooves are equal spacing and specified profile, the latter being both consistent and smooth. Periodic errors in the spacing give rise to 'ghosts', which are satellites of each spectral line

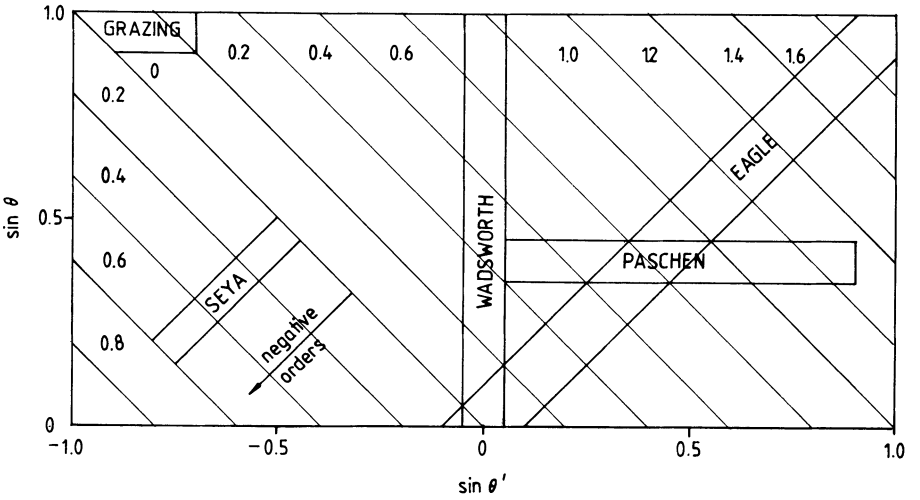

Figure 6.16 Angular ranges covered by various mountings of concave grating. As in Fig. 6.10 the scales along the axes are linear in sin θ and sin θ', respectively, and the diagonals are lines of constant λ/d; for example, the line labelled 0.6 corresponds to $\lambda = 500$ nm for a 1200 line mm^{-1} grating in first order.

at a distance from it inversely proportional to the period of the error – i.e. an error with a period of only a few grooves produces distant ghosts. Very slow variations produce wings to the lines that are largely responsible for the difference between theoretical and actual resolution. Random errors in the spacing produce a general background intensity between the spectral lines, known as 'grass'. The groove shape controls the distribution of intensity among the diffracted orders, and any roughness or random errors in it contribute to the grass.

Once a good master grating has been ruled, replicas of it can be made with epoxy-resin castings that are aluminized after separation from the master. The process can be repeated, replicating the original replicas, for several generations. The quality of the replicas may actually be better than that of the master, because the burrs at the edges of the rulings are transferred to the bottoms of the grooves, but the main purpose of replication is to spread the high cost of ruling a good grating.

Modern gratings are usually blazed, which means that the groove is shaped to concentrate a large fraction of the incident intensity into diffraction at a particular angle. The principle is illustrated in Fig. 6.17: the preferred direction is that corresponding to specular reflection from the 'step' of each groove. If the step is inclined at angle α to the surface of the grating, and N and N′ are the normals to the grating and the step respectively, then

$$\theta - \alpha = -\theta' + \alpha$$

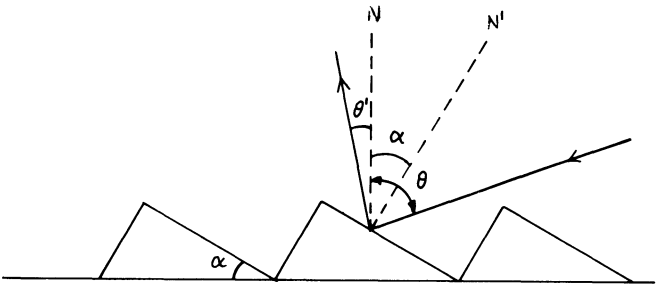

Figure 6.17 Geometry of blazed grating. N and N' are the normals to the grating and the step, respectively, and the blaze wavelength corresponds to specular reflection from the step (see text).

(θ' is negative because it is on the opposite side of N to θ). Therefore,

$$\alpha = (\theta + \theta')/2$$

Combining this with equation 6.1,

$$(m\lambda)_{\text{blaze}} = 2d \sin\alpha \cos(\theta - \alpha)$$

In the Littrow configuration $\theta = \theta' = \alpha$, so the steps are perpendicular to the axis of the spectrometer and the wavelength of maximum efficiency is given by

$$\lambda_b = \frac{2d \sin\theta}{m} = \frac{2d \sin\alpha}{m} \tag{6.17}$$

λ_b for $m = 1$ is normally defined as the blaze wavelength, and it is obviously fixed by the grating ruling; however, the efficiency is maintained over an appreciable range on each side of the optimum, falling to half maximum at about $\frac{2}{3}\lambda_b$ on one side and $\frac{3}{2}\lambda_b$ on the other, and the grating is, of course, also blazed for the second order of $\lambda_b/2$, etc. It is possible to achieve a peak efficiency of about 80%.

The simple macroscopic theory of blaze efficiency breaks down when the groove spacing becomes close to the wavelength: $d \leqslant 3\lambda$. There is a marked difference in efficiency for the two directions of polarization along and perpendicular to the grooves, the latter deviating most from simple theory. This is understandable in a microscopic picture, since the oscillations of the conduction electrons in the direction across the grooves are affected by finite groove width. In fact, abrupt changes in efficiency with λ were first observed by Wood in 1902 and are called after him 'Wood's anomalies'. It is only recently that the problem of relating efficiency to groove shape, size and material has been successfully treated theoretically in terms of collective oscillations of the conduction electrons known as surface plasmons [6].

The most serious problems of ruling a large grating with fine pitch are the wear on the ruling diamond and the control of environmental conditions over a

period of several days. Both of these are avoided with interference ('holographic') gratings. The notion of generating a grating by optical interference was first put forward by Michelson, but its practical implementation had to await the advent of the laser. Such gratings have now been in use for some 20 years and are still being improved. The name 'holographic' presumably arose because the optical arrangement for generating the fringes is similar to that for producing holograms, but 'interference' is a more accurate name.

The gratings are produced by exposing a glass blank coated with photo-resist to the interference pattern produced by two plane wavefronts crossing at an angle, as shown in Fig. 6.18. The interfering wavefronts are derived from the same source, usually the blue line of an argon ion laser, via a beam splitter. Exposure to light changes the solubility of the photo-resist, and subsequent development leaves a surface following the sinusoidal variation of intensity of the interference fringes. From Fig. 6.18 it can be seen that the fringe spacing x from two beams intersecting at an angle 2φ is given by $x=\lambda/(2\sin\varphi)$. The spacing of the grating grooves can be adjusted by inclining the blank at an angle ψ to the plane perpendicular to the fringes, giving

$$d=\frac{x}{\cos\psi}=\frac{\lambda}{2\sin\varphi\cos\psi}$$

Fine spacing requires $\psi=0$ and the largest possible value of φ, about 60° in practice, giving a minimum value of about 0.6λ for d. For the argon ion line at 458 nm this amounts to 3500 lines mm^{-1}, but various tricks can be used to obtain up to 6000 lines mm^{-1}.

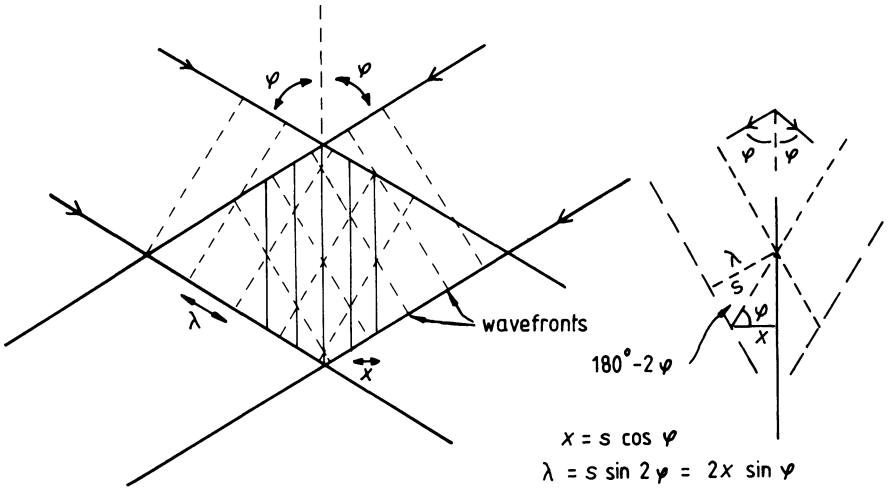

Figure 6.18 Production of interference gratings. The two plane wavefronts crossing at an angle 2φ produce interference maxima spaced by $x=\lambda/2\sin\varphi$, as shown by the vertical lines in the diagram.

Interference gratings have several distinct advantages: the pitch is easily variable and can be finer than that of any ruled grating; diamond wear is avoided; and ghosts and grass should both be absent, though it is necessary to be very careful about stray light and spurious interference patterns during production of the gratings. Their principal disadvantage is the generally lower efficiency, which is a result of the basically sinusoidal groove profile. Considerable effort has been put into generating a sawtooth blaze profile, either during exposure or by subsequent ion etching, and in certain circumstances the efficiency can match that of a blazed ruled grating. Reference [6] should be consulted for further details of both types of grating.

6.7 Gratings for special purposes

6.7.1 VERY HIGH RESOLUTION: ECHELLE GRATINGS

The figure of merit $\mathscr{R}T$ for a grating was shown at the end of section 6.3 to be proportional to the product of order number m and total groove length NH. Since wear on the ruling diamond sets a limit to NH, it is reasonable to look to high orders for high performance. We also noted in relation to equation 6.8, $\mathscr{R} = Wm/d$, that high resolution can be obtained from a coarse ruling in a high order. With $m = 100$ one can get, in round numbers, a resolving power of a million from $W = 100$ mm, and $(1/d) = 100$ lines mm^{-1}. The relatively large value of d allows accurate control of the groove angle so that a large fraction of the intensity is thrown into high orders.

Figure 6.19 is Fig. 6.17 redrawn to bring out the characteristics of echelles. Since they are always used in the Littrow configuration with the light incident on the narrow side of the step, the groove angle α is related to the angle of incidence by $\alpha = (90 - \theta)°$, and this also defines the angle made by the plane of the

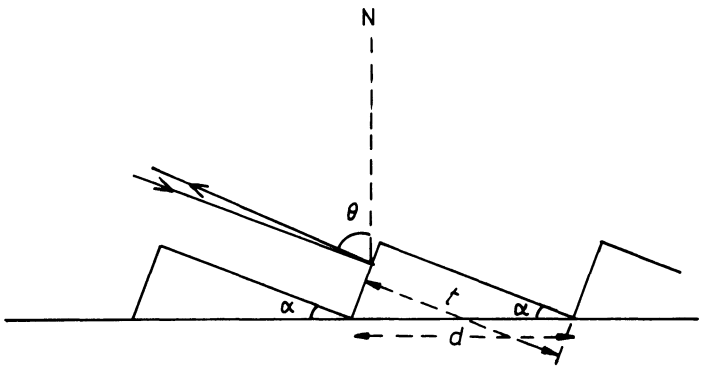

Figure 6.19 Geometry of echelle grating. The blaze angle corresponds to specular reflection from the narrow side of the step, giving $m\lambda = 2d \sin\theta = 2d \cos\alpha = 2t$.

grating with the spectrometer axis. The blaze condition, equation 6.17, becomes for this geometry

$$m\lambda = 2d \sin\theta = 2d\cos\alpha \tag{6.18}$$

Since m is large, this equation can be satisfied by a whole set of wavelengths with neighbouring values of m. An echelle is therefore specified by its blaze *angle* rather than blaze wavelength, and equation 6.18 simply tells us the order number for any particular wavelength. In practice blaze angles are in the range 50–70°. From equation 6.18 and the figure it can be seen that the grating equation can be written as

$$m\lambda = 2t \tag{6.19}$$

where t is the long side of the step. The echelle is beginning to look like an interferometer with a plate separation of t.

As a consequence of the high order number, the order overlap problem is very severe. If the mth order of λ coincides with the $(m-1)$th order of $\lambda+\Delta\lambda$, the wavelength band $\Delta\lambda$ is known as the *free spectral range*. For large m it is found by differentiating the grating equation and setting $\Delta m = 1$:

$$\frac{\Delta\lambda}{\lambda} = -\frac{\Delta m}{m} \Rightarrow |\Delta\lambda| = \frac{\lambda}{m}$$

For $m \sim 100$, $\Delta\lambda$ is only a few nm. The similarity to an interferometer is brought out by expressing the free spectral range in wavenumbers:

$$\Delta\sigma = \frac{|\Delta\lambda|}{\lambda^2} = \frac{1}{m\lambda} = \frac{1}{2t} \tag{6.20}$$

This is identical with the free spectral range of a Fabry–Perot interferometer of plate separation t. The overlap problem is dealt with by introducing cross-dispersion in the form of a small prism or low-resolution grating orientated with its dispersion perpendicular to that of the echelle. Successive orders are then displaced vertically, and the spectrum takes the form of a square array crossed by sloping bands, as indicated in Fig. 6.20. This is obviously a very compact way

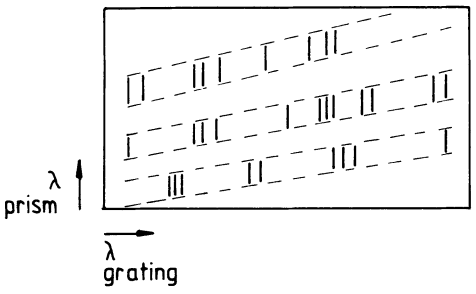

Figure 6.20 Echelle spectrum. The horizontal echelle dispersion is crossed by vertical dispersion from a prism or small grating to avoid order overlap.

of stacking up a great deal of spectral information for detection either photographically or with a two-dimensional array detector, but it is not ideal for photoelectric scanning.

6.7.2 GRATINGS IN THE INFRA-RED

From the grating equation we see that the maximum wavelength for which a given grating can be used is less than $2d$ ($m=1$; $\theta, \theta' \rightarrow 90°$). Gratings for the infra-red therefore require coarser rulings. Moreover, it is usually desirable to work in the first order because of the difficulty of filtering out overlapping orders, and five or six gratings of different pitch are needed to cover the whole infra-red. These blazed gratings of relatively coarse pitch are known as echelettes; echelettes were indeed the first blazed gratings produced. Reflectivities in the infra-red are so high that extra reflections cost almost nothing, and infra-red spectrometers generally use plane gratings in an Ebert–Fastie or Czerny–Turner mounting, taking advantage of the stigmatic imaging and the possibility of using long curved slits to maximize detected energy. Obviously the tolerances on the actual gratings get larger as the wavelength increases, and gratings suitable for the far infra-red can be produced in an ordinary mechanical workshop. Resolving power in this region is in any case usually slit- rather than grating-limited because of the low radiance of the sources available.

The elimination of unwanted orders is a particular problem in the infra-red. An echelette blazed for first order of say 2 μm is also blazed for second order of 1 μm and third order of 500 nm, and the shorter wavelengths may be nearer the peak of the distribution of a thermal source. The problem gets even worse at longer wavelengths. Ideal filters, cutting off all wavelengths shorter than half the wavelength of interest, can rarely be found in practice. Filters relying on selective absorption, selective reflection ('reststrahlen') and selective scattering (Christiansen filters) have all been used for different spectral regions; details may be found in [9, 10].

Two other practical points affect the design of infra-red spectrometers. Atmospheric air shows strong absorption bands from water vapour and carbon dioxide, starting between 1 and 2 μm, and it is necessary either to evacuate the spectrometer or to operate it in a dry gas. Secondly, as seen in Chapter 4, it is usually necessary to provide for cooling the detector.

6.7.3 GRATINGS IN THE FAR ULTRA-VIOLET

Spectroscopic techniques at short wavelengths are governed by the necessity of working in a vacuum from 200 nm to the soft X-ray region, by the lack of transmitting materials below 110 nm (the quartz limit is $\simeq 170$ nm) and by the decreasing reflectivities of mirrors and gratings. Although interferometers and echelle gratings can be used for a little way into the vacuum ultra-violet (VUV), the concave grating reigns supreme through most of the region, for it requires

only one reflection, that from the grating itself. Below about 250 nm dielectric coatings, giving close to 100% reflectivity in the visible and near ultra-violet, have to be replaced by aluminium, or aluminium with a magnesium fluoride overcoating, and the reflectivity drops with wavelength from about 90% to perhaps 70% near 100 nm, below which other metals perform better. Gold, platinum and osmium have all been used below 100 nm, but reflectivities rarely exceed about 20%.

Until recently high-quality ruled gratings of 1200 or 2400 lines mm^{-1} were used for the highest resolution. The theoretical resolution is seldom attained because ruling errors and surface defects all become relatively more important at short wavelengths and detract from performance. Interference gratings should offer significant advantages in this respect, and they are being increasingly used at short wavelengths. Further, it is actually possible to control the focal properties of the grating and to correct some of the aberrations, notably astigmatism, by appropriate shaping of the wavefronts of the interfering beams used to produce the grating.

Below about 30 nm reflectivities are so low that it becomes necessary to use gratings at grazing incidence. The refractive index of any material is less than unity at very short wavelengths, so radiation incident at a sufficiently small angle to the surface is totally *externally* reflected. In terms of the grazing angle φ (the complement of the angle of incidence) the condition for total reflection is

$$\varphi < \varphi_c \quad \text{where} \quad \varphi_c = \lambda\left(\frac{e^2 N_e}{(4\pi\varepsilon_0)\pi m c^2}\right)^{1/2} \tag{6.21}$$

where e and m are the charge and mass of the electron and N_e is the number of electrons per cubic metre in the surface material. Looked at the other way round, this expression gives, for a given angle of incidence, the minimum wavelength for which there is total reflection. Putting in the constants we have

$$(\lambda_{min}) \quad \text{nm} = 3.3 \times 10^{16} N_e^{-1/2} \varphi$$

For $\varphi = 1°$, this gives $\lambda_{min} \simeq 0.66$ nm for aluminium or glass and about 0.25 nm for a heavy metal. To get down to 0.1 nm requires a grazing angle of about 20′. In practice these limits are always affected adversely by surface imperfections, and λ (nm) $\geqslant \varphi$ (degrees) is a good rule of thumb.

One effect of working at grazing incidence is that the grating is effectively foreshortened by the factor $\sin\varphi$. For $\varphi = 1°$, this factor of about 1/50 means that a grating ruled with 1200 lines mm^{-1} is equivalent to a 60 000 line mm^{-1} grating at normal incidence and can still be used in low orders (Fig. 6.21). The smaller effective width of the grating reduces its resolving power; but the resolution of a grazing incidence spectrometer is in practice slit-limited anyway: it is impossible to make slits narrow enough to reach the diffraction limit, and the radiant flux would almost certainly be too small to use them.

Great care has to be taken over surface polish and smoothness in order to reduce surface scattering at these short wavelengths. This can be better done

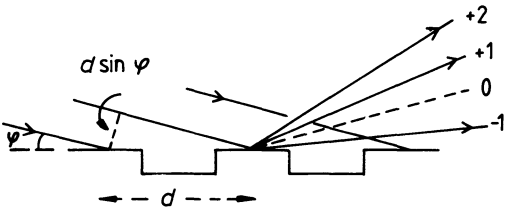

Figure 6.21 Grating at grazing incidence. The effective spacing of the rulings is $d \sin \varphi$. The order numbers are shown on the right.

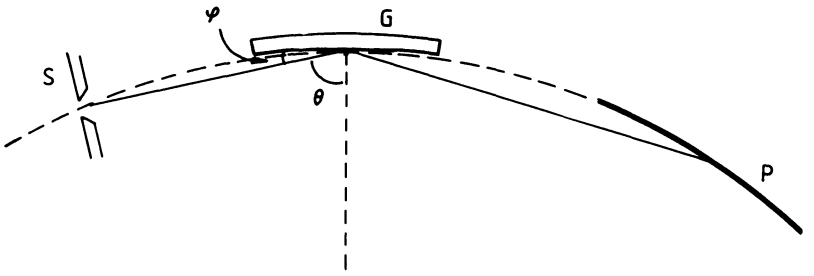

Figure 6.22 Grazing incidence mounting of concave grating. The broken arc is part of the Rowland circle. The grazing angle φ is the complement of the angle of incidence θ.

with laminar than with blazed grooves. These have a shallow square profile, as shown in Fig. 6.21, and actually come rather close to the original bar-and-space model discarded earlier in the chapter.

The grazing incidence mounting is shown in Fig. 6.22. The slit, grating and plateholder all lie on the Rowland circle, but since only a small segment of the circle is used the instrumet is actually quite small, even for a 2 m radius. The mounting and its adjustments have to be carefully designed and machined with great accuracy because such fine angles are involved, and the astigmatism from a standard concave grating is enormous. This defect, with its accompanying light loss, can be largely overcome if the grating is formed on a toroidal rather than a spherical surface – that is, a surface for which the curvature along the length of the grooves is much greater than that perpendicular to them. Further details are given in [6, 11, 12].

References

GENERAL OPTICAL

*1. Born, M. and Wolf, E. (1965) *Principles of Optics*, 2nd edn, Pergamon Press, Oxford.
*2. Ditchburn, R.W. (1963) *Light*, Blackie, London and (3rd edn, 1976) Academic Press, London.

* Further reading, not cited in text.

PRACTICAL USE, MOUNTINGS, ETC.

3. Bousquet, P. (1971) *Spectroscopy and its Instrumentation*, Hilger, London.
4. Harrison, G.R., Lord, R.C. and Loofbourow, J.R. (1948) *Practical Spectroscopy*, Prentice-Hall, Englewood Cliffs, New Jersey.
5. Sawyer, R.A. (1961) *Experimental Spectroscopy*, Dover, New York.

THEORY AND PRACTICE OF GRATING PRODUCTION

6. Hutley, M.C. (1982) *Diffraction Gratings*, Academic Press, New York.
7. Loewen, E.G. (1970) Diffraction gratings for spectroscopy, *J. Phys. E*, **3**, 953.
8. Stroke, G.W. (1967) Diffraction gratings, in *Encyclopaedia of Physics*, Vol. 29, Springer, Berlin, p. 426.

GRATINGS IN THE INFRA-RED

9. Chantry, G.W. (1984) *Long-wave Optics* (2 Vols), Academic Press, New York.
10. Martin, A.E. (1966) *Infra-red Instrumentation and Techniques*, Elsevier, Amsterdam.

GRATINGS IN THE ULTRA-VIOLET

11. Hunter, W.R. (1985) Diffraction gratings and mountings for the VUV spectral region, in *Spectrometric Techniques IV* (ed. G.A. Vanasse), Academic Press, New York.
12. Samson, J.A.R. (1967) *Techniques of Vacuum Ultra-violet Spectroscopy*, Wiley, New York.

SEVEN

Interferometers

Different forms of interferometer are now used for a large number of very different purposes – high precision metrology, measurements of refractive index and density, optical testing, surface studies and microscopy, amongst others. Only two types are important for high resolution spectroscopy: the Fabry–Perot and the Michelson. The Fabry–Perot is a multiple-beam interferometer, capable of extremely high resolution from the near infra-red to the quartz ultra-violet. The Michelson is a two-beam interferometer whose present importance comes from its use as a scanning instrument in Fourier transform spectroscopy. Many other types of interferometer in current use are variations of the Michelson, and one of these, the Mach–Zehnder, is used as a refractometer for the measurement of oscillator strengths (section 12.4) and electron densities in plasmas.

7.1 Some basic concepts of interferometric spectroscopy

Both the Fabry–Perot and the Michelson are division-of-amplitude interferometers, which means that the interfering beams are produced by splitting the incident beam at a partially reflecting surface. In the Fabry–Perot a series of such beams is produced by multiple reflections between two parallel plates with partially reflecting coatings. The Michelson uses a single partially reflecting plate to produce two beams, which are reflected back by plane mirrors and recombined at the beamsplitter.

Figure 7.1 shows the optical arrangement for (a) a Michelson and (b) a Fabry–Perot interferometer. In each case light from an extended source behind an entrance aperture A is collimated by the lens L_1 before impinging on the Fabry–Perot plates F or the beamsplitter B of the Michelson. The lens L_2 images the aperture A at A′ in the plane P. Usually the two mirrors M_1 and M_2 of the Michelson are adjusted so that the image M_1' of M_1 in the beamsplitter is parallel to M_2, as shown in Figs 7.1(a) and 7.2(a), with a distance t between them. For normal incidence the optical path difference x between rays reflected from M_2 and M_1' is $2t$. For incidence at angle θ the geometry of Fig. 7.2(a) shows that

$$x = s(1 + \cos 2\theta) \quad \text{where } s = t \sec \theta$$

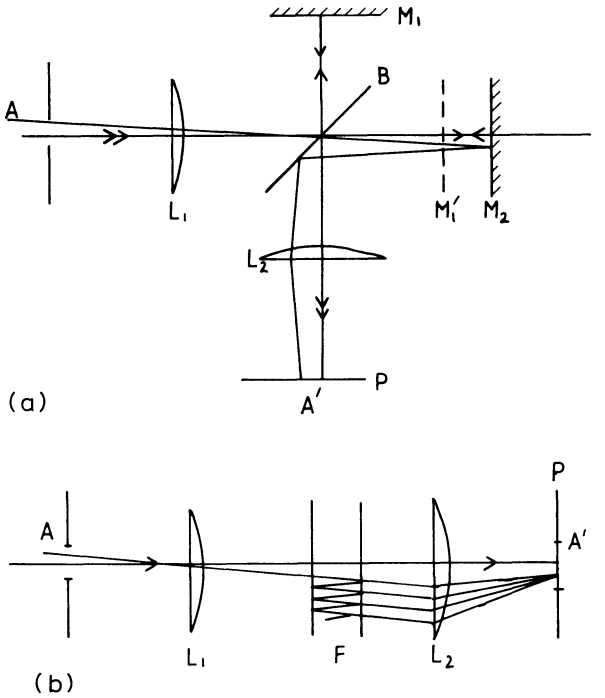

Figure 7.1 Optical arrangements for (a) a Michelson and (b) a Fabry–Perot interferometer. In both diagrams A is the entrance aperture and A′ is its image formed by the collimating and camera lenses L_1 and L_2 in the plane P. In (a) B is the beamsplitter and M_1, M_2 are plane mirrors; M_1' is the image of M_1 in B. In (b) F is the Fabry–Perot interferometer.

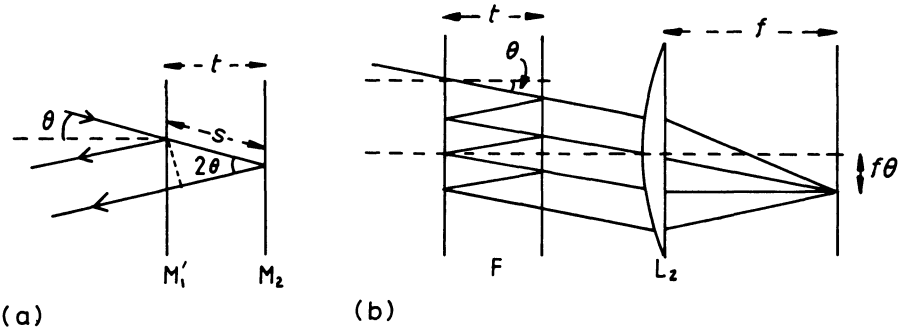

Figure 7.2 Optical path differences for two parallel plates separated by distance t. In (a) the plates are the mirrors M_2 and the image M_1' of M_1, and the optical path difference x between the two reflections is $x=2t\cos\theta$ for angle of incidence θ (see text). In (b) the optical path difference between successive rays in the Fabry–Perot is also $x=2t\cos\theta$, and the rays are focused at a distance $f\theta$ from the optic axis.

Therefore

$$x = \frac{t}{\cos\theta} 2\cos^2\theta = 2t\cos\theta \tag{7.1}$$

Similarly, the optical path difference for each double passage of the Fabry–Perot gap in Fig. 7.2(b) is $2t\cos\theta$. In both types of interferometer the wavefronts interfere constructively when x is an integral number of wavelengths, $m\lambda$, and destructively when x is an odd number of half-wavelengths. The interference fringes are localized in the plane of intersection of the wavefronts, which, since they are parallel, is at infinity. The lens L_2 focuses these fringes in the plane P. For a given wavelength the fringes are loci of constant θ and are known as fringes of equal inclination. It follows from the axial symmetry of the arrangement that the maxima are circles in the plane P of radius $f\theta$ where θ satisfies

$$\left.\begin{aligned} 2t\cos\theta &= m\lambda \\ \text{or} \qquad 2\sigma t\cos\theta &= m \end{aligned}\right\} \tag{7.2}$$

for successive values of m.

Several points about these circular fringes should be noted. First, the order number m decreases outwards from the centre. Secondly, the circles move outwards as t increases (for a given m, larger t requires larger θ); but, thirdly, increasing t crowds the circles more closely together because a given range of θ accommodates more values of m. The number of rings that can be seen at any particular value of t depends on the maximum value of θ, which, with the geometry of Fig. 7.1, is determined by the size of the source or of the aperture A and the focal length of the lens L_1. This lens is *not* essential for the formation of the fringes; its function is to increase the illumination, and in the case of a scanning interferometer it also serves to image the input aperture A in the plane of the fringes.

7.2 Resolution and throughput

There are two ways of using an interferometer as a spectrometer. First, for fixed t equation 7.2 gives λ as a function of θ, and the wavelengths can be measured from the interference fringes photographed in the plane P. Secondly, for fixed θ ($\theta \simeq 0$) equation 7.2 gives λ as a function of t, and wavelengths are found by recording the fringes photoelectrically as t is scanned. The Fabry–Perot may be used either way, but the Michelson is always used in the second mode for reasons that will become apparent later. In this section we consider only the scanning mode, for which the aperture A, and/or another aperture at A′ (Fig. 7.1), ensure that the detector sees only the centre of the ring pattern.

It was shown in section 5.2 that the resolving power of a spectrometer is of order L/λ, where L is the extreme path difference between interfering beams. The

conventional definition of the resolving power in Fourier transform spectroscopy (section 7.7) leads to the expression for the Michelson

$$\mathscr{R}=\frac{2L}{\lambda}=\frac{4t_m}{\lambda} \tag{7.3a}$$

where t_m is the maximum effective separation of the mirrors. For the Fabry–Perot, 'extreme' should refer to the path difference between the first and last of the multiple reflections, but of course there is actually no 'last'–the intensity just drops steadily with the number of reflections. In the next section we define an effective number N_f of interfering beams that depends on the reflectivity of the coatings and is normally in the range 20–50. Then

$$\mathscr{R}\simeq\frac{L}{\lambda}=\frac{2N_f t_m}{\lambda} \tag{7.3b}$$

Thus, in both types of interferometer the resolving power can be increased, in principle without limit, simply by increasing t_m.

The usual price must be paid for high resolution: low throughput. Just as high resolution in a grating spectrometer demands a narrow slit, so in an interferometer it demands a small input aperture. The reason for this can be readily understood from the description of the ring pattern in the previous section. For large t the rings become small. If the aperture is large enough to encompass a complete order–all the way from one maximum to the next of any given wavelength–there is no net change of intensity as t is scanned and no way of distinguishing wavelengths. Suppose wavelength λ has a maximum at the centre, satisfying $m\lambda=2t_m$, while wavelength $\lambda-\delta\lambda$ forms a ring of angular radius θ satisfying

$$m(\lambda-\delta\lambda)=2t_m\cos\theta\simeq 2t_m\left(1-\frac{\theta^2}{2}\right)$$

(θ is always small enough to justify this expansion.) Then

$$m\delta\lambda=t_m\theta^2$$

To resolve λ from $\lambda-\delta\lambda$ we have to exclude this ring from the aperture. This limits the angular aperture to the value θ_m given by

$$\theta_m^2\leqslant\frac{m}{t_m}\delta\lambda=\frac{2\delta\lambda}{\lambda}=\frac{2}{\mathscr{R}}$$

In terms of the radius a of the input aperture and the focal length f of the collimating lens

$$\theta_m\leqslant\frac{a}{f}=\sqrt{\frac{2}{\mathscr{R}}} \tag{7.4}$$

For $\mathscr{R}=2\times10^5$, equivalent to a high resolution grating, $\theta_m\simeq 3$ mrad. For f of order 1 m, a is a circle of radius 3 mm, which is a very large input aperture when

compared with a slit of 10 mm × 10 μm. This is the basis of the throughput advantage, discussed in section 7.9.

At this stage it is more useful to discuss the two types of interferometer separately. We take first the Fabry–Perot, its intensity distribution and the ways in which it is used, and then the Michelson interferometer and the technique of Fourier transform spectroscopy. For fuller treatments see [5, 6] for the Fabry–Perot and [7–9] for Fourier transform spectroscopy.

7.3 Fabry–Perot interferometers: intensity distribution

A Fabry–Perot consists of two glass or fused silica plates, usually 2–10 cm in diameter, with their facing surfaces flat and parallel to better than 1/50 of a wavelength. These surfaces are coated with metal or dielectric films to give a reflectivity which is usually in the range 85–95%. The back surfaces have a wedge angle sufficient to throw reflections from them clear of the main fringe pattern. The plates are frequently held in optical contact with a fixed spacer – this arrangement is known as an etalon – but they may also be mounted on piezoelectric crystals or other devices suitable for making small changes in separation.

From equation 7.1 and Fig. 7.2 the path difference for each double passage of the etalon at angle of incidence θ is $2t\cos\theta$. Allowing for a medium of refractive index n between the plates, the optical path difference is

$$\begin{aligned} x &= 2nt\cos\theta \\ &\simeq 2t\cos\theta \end{aligned} \tag{7.5}$$

Strictly, θ in this equation should be the angle in the medium rather than the angle of incidence, but so long as the medium is a gas the difference is of order $1:10^3$ and is unimportant. For the same reason it is usually adequate to take $n=1$. Although we already know that $x=m\lambda$ defines the fringe maxima, it is necessary to find the intensity distribution in the fringes – i.e. the instrument function, which is quite different from that of a grating – to quantify the properties of the Fabry–Perot as a spectrometer.

Figure 7.3 shows the amplitudes of the interfering beams in terms of the incident amplitude A_0; s and r are the amplitude coefficients of transmission and reflection of each of the films (assumed to be the same for both plates). The first transmitted beam has amplitude A_0s^2, and each double passage introduces a factor r^2 to the amplitude and a phase shift δ given by

$$\left.\begin{aligned} \delta &= 2\pi x/\lambda = 2\pi\sigma x \\ \delta &= 4\pi\sigma t\cos\theta \end{aligned}\right\} \tag{7.6}$$

or

When all these beams are superposed in the focal plane of the lens, the resultant

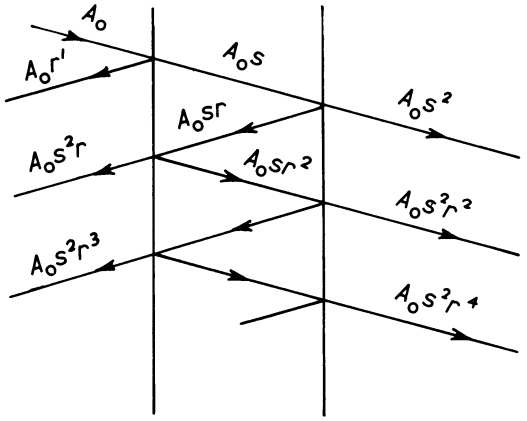

Figure 7.3 The amplitudes of successive transmissions and reflections in a Fabry–Perot interferometer. s and r are the amplitude coefficients of transmission and reflection of the plate coatings. The first backward-going reflection, A_0r', has a phase shift of π with respect to the other beams.

complex amplitude is

$$A = A_0 s^2 (1 + r^2 e^{i\delta} + r^4 e^{2i\delta} + \cdots) = \frac{A_0 s^2}{1 - r^2 e^{i\delta}}$$

To get from complex amplitude to intensity, A is multiplied by its complex conjugate, as done for the grating in section 6.2:

$$I = AA^* = \frac{s^4}{(1 - r^2 e^{i\delta})(1 - r^2 e^{-i\delta})} A_0^2$$

Introducing the intensity coefficients, transmissivity $T = s^2$ and reflectivity $R = r^2$, and the incident intensity $I_0 = A_0^2$, we have

$$I = \frac{T^2}{1 - 2R\cos\delta + R^2} I_0 = \frac{T^2}{(1-R)^2 + 2R(1 - \cos\delta)} I_0$$

$$= \frac{T^2 I_0}{(1-R)^2 + 4R\sin^2(\delta/2)}$$

Therefore,

$$I = I_0 \left(\frac{T}{1-R} \right)^2 \frac{1}{1 + [4R/(1-R)^2]\sin^2(\delta/2)} \tag{7.7}$$

Equation 7.7 is known as the Airy distribution.

The maxima of the distribution are given by

$$\delta/2 = m\pi \tag{7.8}$$

or

$$\sigma x = m$$

which is identical with equation 7.2, $2\sigma t\cos\theta = m$. The peak intensity is

$$I_{max} = I_0\left(\frac{T}{1-R}\right)^2 \tag{7.9}$$

In the ideal case of no absorption, $T+R=1$, and *the peak intensity is equal to the incident intensity.* The minima occur at values of δ half-way between the maxima, when $\delta/2=(m+\frac{1}{2})\pi$, and their intensity is

$$I_{min} = I_0\left(\frac{T}{1+R}\right)^2 \tag{7.10}$$

The *contrast* C is defined by the ratio of maxima to minima:

$$C=\frac{I_{max}}{I_{min}}=\left(\frac{1+R}{1-R}\right)^2 \tag{7.11}$$

For $R\simeq 90\%$, C is a few hundred.

The Airy distribution can conveniently be written in terms of I_{max} as

$$I=\frac{I_{max}}{1+f\sin^2(\delta/2)}=\frac{I_{max}}{1+f\sin^2\pi\sigma x} \tag{7.12}$$

where $$f=\frac{4R}{(1-R)^2}$$

Figure 7.4 shows the distribution plotted as a function of x for two different

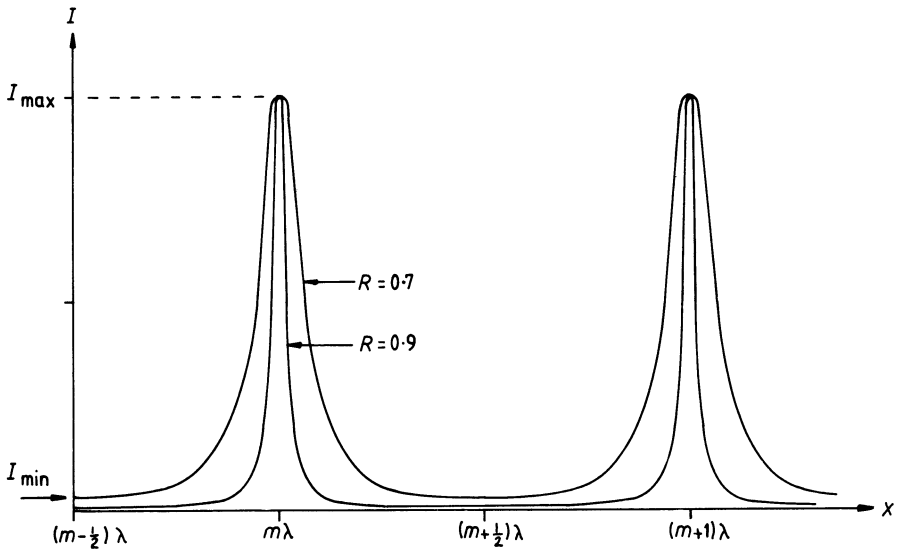

Figure 7.4 Airy distribution as a function of optical path difference x for two different reflectivities, $R=0.7$ and $R=0.9$.

reflectivities, 70% and 90%. The sharpness of the fringes obviously depends critically on R.

By rescaling the abscissae we can equally well regard Fig. 7.4 as a plot of intensity versus wavenumber at fixed x. In this guise it represents the output of the instrument from a monochromatic input – in other words the instrument function. Figure 7.5 shows the distribution for $R = 90\%$ appropriately relabelled. The single monochromatic line has been reproduced at intervals $\Delta\sigma$ with a FWHM (full-width half-maximum) of $\delta\sigma_{1/2}$. The repeat distance $\Delta\sigma$ represents one order of interference and is the *free spectral range.*

We now need expressions for $\Delta\sigma$ and $\delta\sigma_{1/2}$ in terms of the properties of the etalon. $\Delta\sigma$ is easily found by differentiating equation 7.8 for constant x and setting $\Delta m = 1$:

$$\frac{\Delta\sigma}{\sigma} = \frac{\Delta m}{m} \Rightarrow \Delta\sigma = \frac{\sigma}{m} = \frac{1}{x} \simeq \frac{1}{2t} \tag{7.13}$$

This equation is identical to equation 6.20 for the free spectral range of the echelle. However, t, and hence m, is typically from one to three orders of magnitude higher in the Fabry–Perot, and the problem of overlapping orders is correspondingly more severe. For $t = 1$ mm, $\Delta\sigma = 5\ \text{cm}^{-1}$, which is only about 0.1 nm in the visible, and it is quite usual to work at considerably higher values of t than this. Cross-dispersion is nearly always essential to avoid a huge ambiguity problem; this is discussed further in section 7.5.

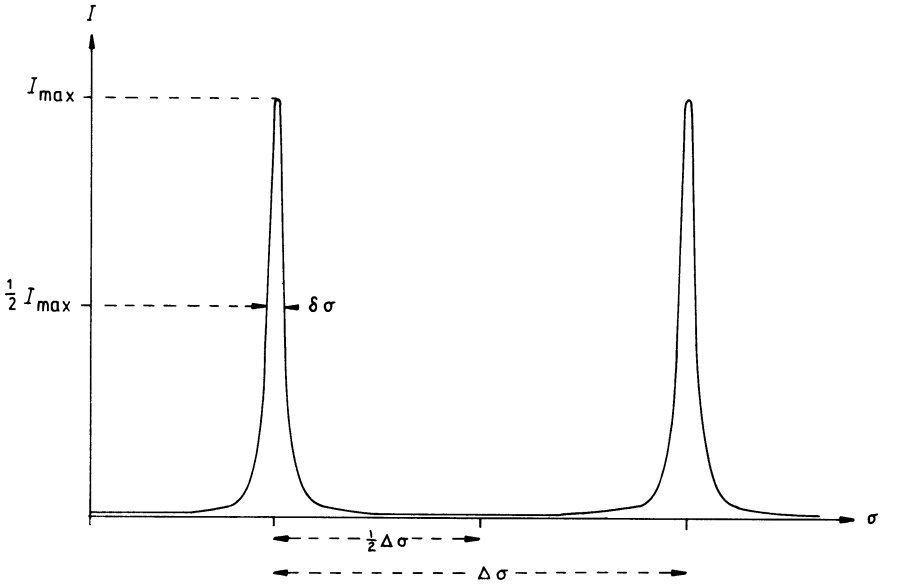

Figure 7.5 Airy distribution as a function of wavenumber σ. $\Delta\sigma$ is the free spectral range. The distribution shown is for $R = 0.9$, and the FWHM $\delta\sigma$ is $(1/28)\Delta\sigma$.

To find the FWHM $\delta\sigma_{1/2}$ of the Airy distribution we need to step off by an amount $\frac{1}{2}\delta\sigma_{1/2}$ either side of the maximum such that I in equation 7.12 is equal to $\frac{1}{2}I_{\max}$:

$$1+f\sin^2\left(\pi\frac{\delta\sigma_{1/2}x}{2}\right)=2 \Rightarrow \sin\left(\pi\frac{\delta\sigma_{1/2}}{2}\right)=\frac{1}{\sqrt{f}}$$

Replacing x by $1/\Delta\sigma$ (equation 7.13),

$$\pi\frac{\delta\sigma_{1/2}x}{2}=\frac{\pi}{2}\frac{\delta\sigma_{1/2}}{\Delta\sigma}$$

$\delta\sigma_{1/2}/\Delta\sigma$ is necessarily small if the fringes are sharp, so we can replace the sine by the angle to give

$$\delta\sigma_{1/2}=\frac{2}{\pi\sqrt{f}}\Delta\sigma=\frac{1-R}{\pi\sqrt{R}}\Delta\sigma \tag{7.14}$$

The ratio of free spectral range to FWHM measures the fine-ness, or 'finesse' of the fringes, and the coefficient of finesse is defined by

$$N_f\equiv\frac{\Delta\sigma}{\delta\sigma_{1/2}}=\frac{\pi\sqrt{f}}{2}=\frac{\pi\sqrt{R}}{1-R} \tag{7.15}$$

It is reasonable to take the FWHM as approximately equal to the resolution limit of the etalon. We show in the next section that for all practical purposes the two are the same. On this basis the resolving power can be written, using equations 7.13 and 7.15, as

$$\mathscr{R}=\frac{\sigma}{\delta\sigma_{1/2}}=\frac{\sigma}{\Delta\sigma}\frac{\Delta\sigma}{\delta\sigma_{1/2}}=mN_f \tag{7.16}$$

and it is clear that the finesse N_f can be interpreted as the effective number of interfering beams. Table 7.1 shows the dependence of N_f on R. Obviously a large N_f is desirable in order to obtain high resolving power without restricting the free spectral range too much, but there are a number of practical considerations (described in the next section) limiting it to about 50.

We conclude this section by attacking a popular misconception about the Fabry–Perot. It is easy to imagine that the transmitted intensity through two plates, each with reflectivity say 90%, should be reduced to 1%. This, of course, is true of the average level, but it is not true of the peaks. Provided there are no

Table 7.1 Finesse N_f of Fabry–Perot for different values of coating reflectivity R

R	0.70	0.75	0.80	0.85	0.90	0.95
N_f	8.3	10.3	13.3	18.3	28.3	58.0

absorption losses, the peak intensity is 100% of the incident intensity. What the etalon does is to redistribute the energy, piling it up in the narrow bright fringes at the expense of the regions between peaks. Further, as indicated in Fig. 7.3, there is another set of interfering beams going back towards the source. The first of these, marked A_0r' in the figure, has a phase shift of π relative to the others if absorption in the plate coatings is negligible. It can be shown, though we do not take the space to do it here, that the intensity distribution produced by these beams is exactly complementary to that described by equation 7.7 – that is, it consists of narrow dark fringes on a bright background. Indeed, this pattern can be deduced without any algebra at all from the principle of conservation of energy.

7.4 Fabry–Perot interferometers: resolution and limitations

We start by justifying the equivalence of resolution limit and FWHM assumed in equation 7.16. Rayleigh's criterion is directly applicable only to a diffraction-limited instrument function, but it can be adapted for other line shapes, as described in section 5.2, by taking a 20% dip between the peaks of two equal lines as the resolution criterion. At first sight the FWHM, $\delta\sigma_{1/2}$, is not a suitable candidate for the resolution limit, since it would appear that the superposed profiles of two equal lines separated by $\delta\sigma_{1/2}$ would show no dip at all at the midpoint where each has intensity $\frac{1}{2}I_{max}$; but the Airy distribution does not fall to zero in the wings of the lines, and as shown in Fig. 7.6 each of the line peaks is raised by the wing of the other to a value I_p, where $I_p = I_{max} + I_w$. I_w is the intensity at a distance $\delta\sigma_{1/2}$ from the maximum and can be found from equations 7.12 and 7.13:

$$I_w = \frac{I_{max}}{1 + f\sin^2\pi\left(\dfrac{\delta\sigma_{1/2}}{\Delta\sigma}\right)}$$

The sine can again be replaced by the angle, and using equation 7.15 we get

$$\sin^2\left(\pi\frac{\delta\sigma_{1/2}}{\Delta\sigma}\right) \simeq \pi^2\left(\frac{\delta\sigma_{1/2}}{\Delta\sigma}\right)^2 = \frac{4}{f}$$

Therefore,

$$I_w = \frac{I_{max}}{1+4} \Rightarrow I_p = \frac{6}{5}I_{max}$$

The intensity in the dip is thefore $\frac{5}{6}$, or 83%, of the peak intensity, which is quite close enough to the value of 81% required by the strict application of Rayleigh's criterion. Thus, for all practical purposes the resolution limit $\delta\sigma$ is identical with

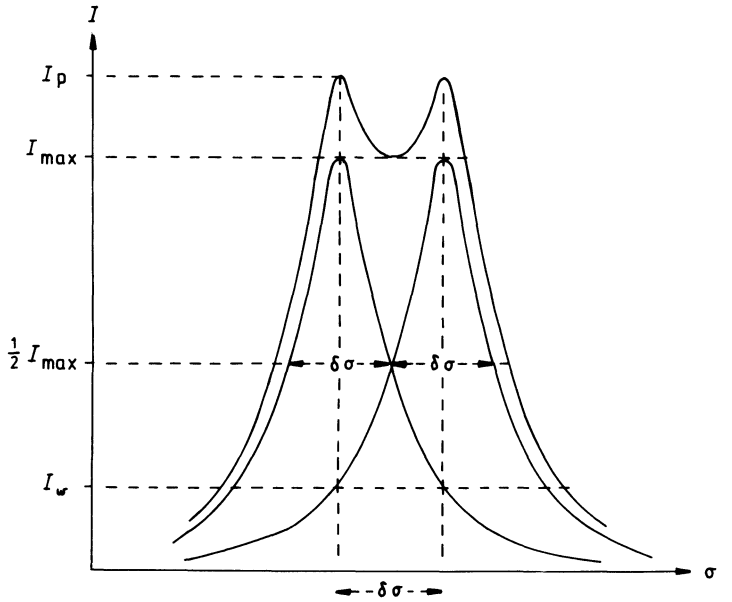

Figure 7.6 Rayleigh's criterion applied to the Airy distribution. The peak intensity I_p is the sum of the maximum from one line, I_{max}, and the intensity I_w in the wing of the other at a distance $\delta\sigma$ from its centre.

the FWHM $\delta\sigma_{1/2}$ and is given by equation 7.15:

$$\delta\sigma = \frac{\Delta\sigma}{N_f}$$

Equation 7.16 for the resolving power is therefore justified:

$$\mathscr{R} \equiv \frac{\sigma}{\delta\sigma} = \frac{\sigma}{\Delta\sigma} N_f = mN_f$$

It is obvious from putting in a few figures that the maximum resolving powers attainable from gratings can easily be exceeded with Fabry–Perots. For $N_f = 25$ a resolving power of a million in the visible requires a plate separation of only 1 cm, and it is not difficult to increase this by an order of magnitude. The Fabry–Perot is, in fact, equivalent to a grating of width $N_f t$, with the additional advantage that the 'width' can be easily varied according to the needs of a particular experiment simply by using different spacers. Since the interferometer also has a substantial advantage over a grating in light throughput (section 7.2), we need to take a look at its limitations to see why it is so much less widely used.

The principal limitation is the small free spectral range: $\Delta\sigma = 1/2t$. For a very sparse line spectrum it may be possible to isolate the relevant spectral ranges with interference filters, but it is nearly always necessary to use an auxiliary prism or grating spectrograph (see next section). To reduce the demands on the cross-dispersing instrument, one would like to use the largest possible N_f with correspondingly small t to attain a given resolution. N_f is limited to 20–50, depending on the wavelength region, for two possible reasons. The first concerns loss of intensity from absorption in the films. Writing $T = 1 - R - A$, where A is the absorption, equation 7.9 becomes

$$I_{max} = I_0\left(1 - \frac{A}{1-R}\right)^2$$

The peak intensity is close to I_0 only if A is small compared with $1-R$, a condition that obviously becomes harder to meet as R is increased. Multilayer dielectric films are, in fact, so efficient in the visible and near infra-red that absorption is not usually a serious limitation, but very high reflectivities can be obtained only over a rather narrow spectral band near the wavelength for which the film thicknesses have been optimized, so that a compromise is usually necessary. In the ultra-violet region the materials of high refractive index in the multilayer coatings tend to start absorbing. Reflectivities of about 90% can be attained down to about 250 nm, but at shorter wavelengths aluminium films have to be used and 80–85% is more likely. Fabry–Perots of low finesse have been used down to about 140 nm, but for most purposes they are not useful much below 200 nm.

The second limitation on finesse is set by the imperfections of the actual interferometer plates. The effect of any departure from complete flatness is that all the rays emerging from the interferometer at a given angle have varying phase differences. A bump of $\lambda/50$ changes the local path difference by $\lambda/25$ and hence changes the phase by 1/25 of an order. If the fringe maximum can wander about through 1/25 of an order, the finesse cannot be greater than 25. Although interferometer plates have been worked to $\lambda/100$, or even better over a small area, $\lambda/50$ is a more usual figure – and by λ we mean visible light, by which this very skilled work is monitored. The plates are relatively 'rougher' in the ultra-violet. Little is gained by using a reflection finesse appreciably greater than the plate finesse, and a detailed treatment shows that the two should be approximately matched [1, 5]. The effective finesse is then 0.6 of either, usually 25–30 in the visible but perhaps higher in the infra-red.

A disadvantage of quite a different kind arises from the non-zero minimum in the Airy intensity distribution. This tends to mask weak lines with strong neighbours. For the same reason the Fabry–Perot is not ideal for absorption spectroscopy: the absorption lines are to some extent filled in by the minima from the surrounding continuum.

Finally, Fabry–Perot interferometers are much more limited in spectral range than gratings because they require transmitting and very flat substrates for the

partially reflecting coatings. We have seen that their effective short-wavelength limit is close to 200 nm. The infra-red is less of a problem, even beyond the quartz cut-off at a few μm, because the flatness tolerances become easier to meet as the wavelength increases. Fabry–Perots with stretched polymeric films ('melinex') have been used above the limits of the alkali halide crystals, and indeed metallic mesh 'plates' have been taken right into the sub-millimetre range. However, the main competitor with grating spectrometers in the infra-red is Fourier transform rather than Fabry–Perot interferometry.

7.5 Fabry–Perot interferometers: methods of use

The traditional way of using the Fabry–Perot is to photograph the ring pattern and measure the ring diameters with a travelling microscope to obtain wavelengths [6]. Nowadays Fabry–Perots are normally used in the photoelectric scanning mode. Either way, an auxiliary spectrograph or monochromator is nearly always necessary because of the small free spectral range. Prism instruments usually provide adequate resolution.

For photographic use the spectrograph must be stigmatic. The interferometer can be mounted either outside the spectrograph, the fringes being focused on the slit with a lens of high quality, or inside it in the parallel beam between the collimator and the prism (Littrow instruments are *not* suitable for internal mounting). The slit image defines a thin rectangular section through the ring system, and the interference fringes are imaged as short arcs in the focal plane of the spectrograph, as shown in Fig. 7.7. The width of the slit has to match the free spectral range of the interferometer unless the line(s) under investigation are well isolated from other lines in the spectrum. The complete spectrum forms a two-dimensional array a little like that of the echelle, except that in this case the high resolution runs along the length of the slit and the low resolution dispersion is perpendicular to the slit. The wavenumber scale is quadratic in fringe diameter D, as can be seen from equation 7.2: within a given order m,

$$\Delta\lambda \equiv \lambda_1 - \lambda_2 = \frac{2t}{m}(\cos\theta_1 - \cos\theta_2) = \frac{2t}{m}\left(\frac{\theta_2^2}{2} - \frac{\theta_1^2}{2}\right)$$

$$= \frac{\lambda}{8f_2^2}(D_2^2 - D_1^2)$$

where f_2 is the focal length of the fringe-forming lens. The scaling factor λ/f_2^2 can be found by measuring diameters for successive orders of one wavelength. Wavelength measurements made in this way can be extremely accurate, but intensity measurements suffer from the usual photographic difficulties of plate calibration and small dynamic range.

The photoelectric scanning mode requires a circular aperture to isolate the centre of the ring pattern, as described in section 7.2. Ideally, to exploit fully the

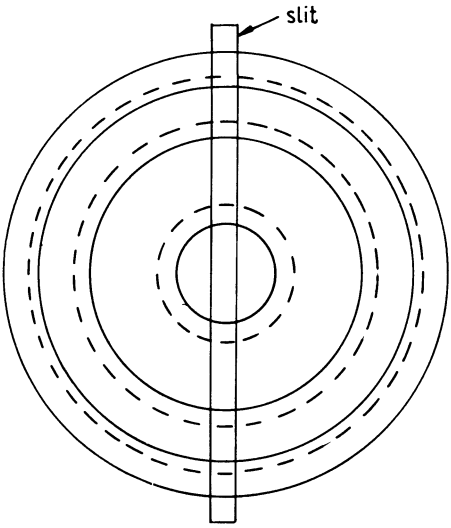

Figure 7.7 Fabry–Perot ring pattern crossed with spectrograph slit. The full and broken circles are maxima for two different wavelengths, λ_1 and λ_2.

potentially high throughput the slit width of the auxiliary spectrometer should match the diameter of the aperture. Since the latter decreases only as the inverse square root of the resolving power (equation 7.4), whereas the free spectral range decreases as $1/\mathscr{R}$ for a given finesse (equation 7.16), a combination of high resolution and a crowded spectrum may require for the match an impracticably large linear dispersion.

The actual scan can be very short. In principle all the available information can be obtained from a single order, although it is obviously desirable to scan over several orders. Even so, since one order represents a path change of $\lambda/2$, a scan of a few μm is normally adequate. It is necessary to keep the plates parallel to about $\lambda/50$ during the scan. This can be done by physically moving one plate with a piezoelectric, magnetic or spring-strip device, but it can also be done by enclosing an etalon with a fixed spacer in an airtight box and changing the pressure. Using equation 7.5 in the form $x = 2nt$, it can be seen that a scan of one order, $\Delta x = \lambda$, requires

$$\Delta n = \frac{\lambda}{2t} = \frac{1}{m} = \frac{N_f}{\mathscr{R}}$$

so that for a finesse of 30 and a resolving power of a million $\Delta n = 3 \times 10^{-5}$, which amounts to a change in pressure of 0.1 atm. The wavelength scale is effectively linear in either n or t:

$$\Delta\lambda = \Delta x/m$$

where

$$\Delta x = t\Delta n \quad \text{or} \quad \Delta x = n\Delta t$$

and the change in m is usually less than $1:10^4$.

The problem of limited spectral range is sometimes tackled by using two, or even more, etalons in series. If one spacing is an exact multple (5, say) of the other, the maxima of the two etalons coincide except that the longer etalon has five maxima for every one of the shorter. The latter acts as a filter, transmitting every fifth order of the longer, high-resolution etalon. The fine tuning necessary to obtain the exact ratio can be made by adjusting the pressure in one of the etalons. The advantage of the double etalon is that the intensity in the transmitted peaks is reduced scarcely at all, provided the absorption in the coatings is small, whereas an auxiliary spectrometer is likely to give at best 50% efficiency even if it does not cut the throughput. The filter etalon may solve a problem sometimes met in studying Zeeman or hyperfine structure, where the high resolving power required of the main etalon to separate the components may spread the whole pattern over a wavelength range exceeding its free spectral range; but the intensity away from the maxima of the filter etalon does not drop to zero, so the 'suppressed orders' are not entirely suppressed and have to be allowed for in interpreting the spectrum.

Before leaving the Fabry–Perot interferometer, we should mention its use in the determination of standards. The first such application was the historic determination of the absolute wavelength of the red line of cadmium in terms of the International Metre in Paris in 1905 by Fabry, Perot and Benoit. When the international standard of length became the orange krypton line, the same type of measurement was reinterpreted as a measurement of the metre in terms of wavelength. The present importance of the Fabry–Perot in this field is in the establishment of secondary and tertiary wavelength standards.

7.6 Michelson interferometers and Fourier transform spectroscopy

The Michelson interferometer set up as in Figs 7.1(a) and 7.2(a) gives fringes of equal inclination. Regarded simply as a Fabry–Perot interferometer with a finesse of 2, it obviously has no merit. Its value as a spectrometer is due to the fact that the output signal as a function of the path difference x between the two beams – the interferogram – is the Fourier transform of the spectrum; the spectrum can therefore be recovered by taking the inverse Fourier transform of the interferogram. This important relation between the interferogram and the spectrum should be examined before going into the practical implementation of Fourier transform spectroscopy.

The complex amplitude from the superposition of two coherent beams of amplitude a_1 and a_2 and wavenumber σ with an optical path difference x is given by

$$A = a_1 + a_2 e^{i\delta} \quad \text{where} \quad \delta = 2\pi\sigma x$$

Multiplying by the complex conjugate,

$$I_\sigma = AA^* = a_1^2 + a_2^2 + a_1 a_2 (e^{i\delta} + e^{-i\delta})$$
$$= a_1^2 + a_2^2 + 2a_1 a_2 \cos\delta \tag{7.17}$$

If a_1 and a_2 both have the same value a,

$$I_\sigma(x) = 2a^2(1 + \cos\delta) = \bar{I}_\sigma(1 + \cos 2\pi\sigma x) \tag{7.18}$$

I_σ measures the flux, in watts or photons s^{-1}, falling on the detector, $\bar{I}_\sigma$ is its mean value, and $x = 2t\cos\theta$ (equation 7.1). The interferometer is used in the scanning mode: the centre of the ring pattern is isolated by an aperture ($\theta \simeq 0$) and x is changed by moving one (or both) of the mirrors, as described in section 7.2. The signal then varies sinusoidally between the limits $2\bar{I}_\sigma$ and 0, as illustrated in Fig. 7.8. We get from monochromatic radiation to a real light source by summing the contributions from all spectral elements to I_σ as there is no coherence between them. If the source has spectral radiance $L(\sigma)$, the average flux in the range $d\sigma$ reaching the detector is proportional to $L(\sigma)d\sigma$ and to some slowly varying function of σ which describes the spectral dependence of the transmission and reflection coefficients. To conform with the usual notation, this modified source function is designated $B(\sigma)$; its units are W (or photons s^{-1}) per cm^{-1}. Equation 7.18 then becomes

$$I_0(x) = \int_0^\infty B(\sigma)(1 + \cos 2\pi\sigma x)\,d\sigma$$
$$= \int_0^\infty B(\sigma)\,d\sigma + \int_0^\infty B(\sigma)\cos 2\pi\sigma x\,d\sigma$$
$$= \bar{I} + \int_0^\infty B(\sigma)\cos 2\pi\sigma x\,d\sigma \tag{7.19}$$

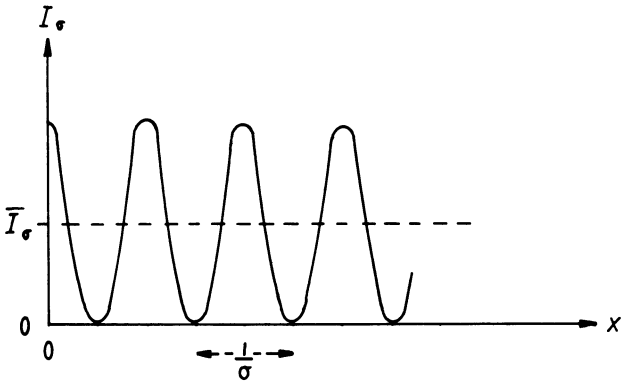

Figure 7.8 The interferogram for a monochromatic source of wavenumber σ: $I(x) = \bar{I}_\sigma(1 + \cos 2\pi\sigma x)$.

The first term in equation 7.19 is the constant level in the absence of modulation, and the second term contains all the information on the spectrum. The lower limit of integration of this term can be taken as $-\infty$ instead of zero because $B(\sigma) = 0$ for negative σ. We define $I(x)$ as the modulated part of the interferogram only, and Fig. 7.9 is an example of what it might look like

$$I(x) = I_0(x) - \bar{I}$$

Therefore

$$I(x) = \int_{-\infty}^{\infty} B(\sigma) \cos 2\pi\sigma x \, d\sigma \tag{7.20}$$

$I(x)$ is thus the cosine transform of $B(\sigma)$, and $B(\sigma)$ can be recovered from it by the inverse cosine transform. The Fourier transform pair is

$$\left.\begin{aligned} I(x) &= \int_{-\infty}^{\infty} B(\sigma) \cos 2\pi\sigma x \, d\sigma \quad \text{(W)} \\ B(\sigma) &= \int_{-\infty}^{\infty} I(x) \cos 2\pi\sigma x \, dx \quad \text{(W per cm}^{-1}\text{)} \end{aligned}\right\} \tag{7.21}$$

Looked at from a slightly different point of view, the Michelson interferometer imposes a characteristic modulation on each optical frequency which enables the signal to be subsequently unscrambled. If one of the mirrors is moved at a speed $v/2$, the path difference changes at a rate v; $x = vt$, and equation 7.20 becomes

$$I(t) = \int_{-\infty}^{\infty} B(f) \cos 2\pi f t \, df$$

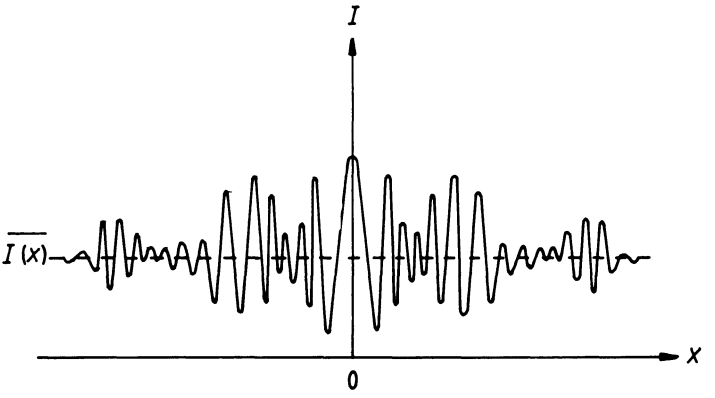

Figure 7.9 Schematic interferogram for a polychromatic source $B(\sigma)$: $I(x) = \bar{I} + \int_0^{\infty} B(\sigma) \cos 2\pi\sigma x \, d\sigma$.

where the frequency f is given by

$$f = v\sigma \tag{7.22}$$

f is in the audiofrequency region; for a wavelength of 1 μm and $v = 1$ mm s^{-1}, for example, $f = 1$ kHz. The detector receives signals simultaneously from every spectral element in the source, each encoded with a modulation frequency proportional to its optical frequency.

This form of spectroscopy is given the name 'multiplex', by contrast with the dispersive techniques in which the different spectral elements are spread out spatially. The recovery of the spectrum from the interferogram is considered in a little more detail in the next section. We shall look first at some important aspects of the generation of the interferogram.

It was assumed following equation 7.17 that the two interfering beams both had the same amplitude. If this is not the case, the depth of modulation decreases, but the mean value stays the same. This is undesirable because the modulated part of the signal contains all the useful information and the constant part contributes only noise. As normally arranged, each beam has one reflection and one transmission at the partially reflecting coating of the beamsplitter, so equal amplitudes are achieved regardless of the actual value of the reflectivity. The flux I_0 is proportional to the product RT of the reflection and transmission coefficients, and the mean value $\bar{I}$ has its maximum value of $2RT$ (50% if there is no absorption) when $R = T$. The other half of the incident energy goes back towards the source in the complementary interference pattern. The amplitudes of the interfering beams in this output are not balanced unless $R = T$.

The next point concerns the matching of the optical paths in the two arms of the interferometer. Figure 7.10 shows that one beam has one passage through the beamsplitter plate while the other has three. The additional optical path can not be compensated at all wavelengths by increasing the air (or vacuum) path in

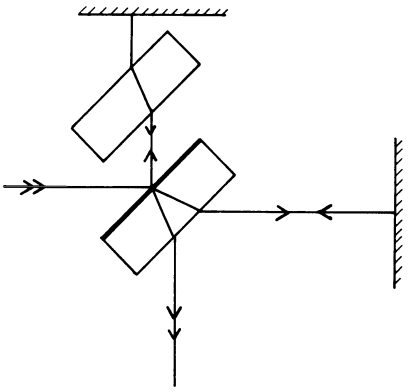

Figure 7.10 Compensating plate used to match optical paths through beamsplitter material.

the other beam because of the dispersion of the beamsplitter material. The compensating plate shown in the figure is introduced to equate the optical paths, but if the thicknesses are not exactly matched the interferogram is to some extent asymmetric about zero path difference, as indicated in Fig. 7.11.

The optical tolerances for a two-beam interferometer are normally specified as $\pm\lambda/4$. This sounds an easy requirement to meet after the $\lambda/50$ of the Fabry–Perot, but it has to be maintained over much longer scans of the moving mirror (see next section)—possibly tens of centimetres instead of a few wavelengths. The tolerances for both optical and mechanical precision get increasingly difficult to meet towards shorter wavelengths, and this is certainly one of the principal reasons why Fourier transform spectroscopy, though a well-established technique in the infra-red, has only rather recently been extended to the visible and ultra-violet. One way of easing the demands on the scanning system is to use retroreflectors (cube corners or catseyes) instead of plane mirrors (Fig. 7.12). These are tilt-invariant, and they have the additional

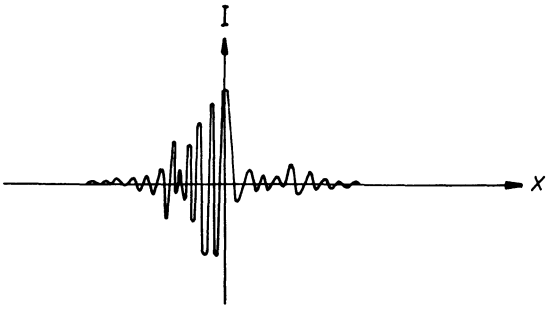

Figure 7.11 Asymmetric interferogram from unmatched optical paths.

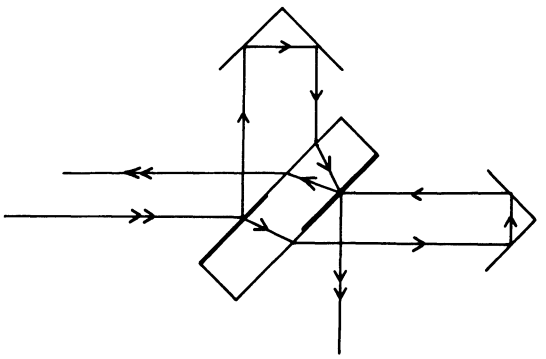

Figure 7.12 Retro-reflectors used instead of plane mirrors to make Michelson interferometer tilt-invariant.

advantage that the complementary interferogram returning towards the source is laterally displaced from the input beam and can be picked up by a second detector for the improvement of the signal-to-noise ratio. The arrangement shown in the figure, with the two halves of the beamsplitter coated on opposite sides, also dispenses with the need of a separate compensating plate.

7.7 Fourier transform spectroscopy: instrument function, resolution, sampling and aliasing

The ideal transform pair, equation 7.21, is never realized in practice. Quite apart from any imperfections in the interferometer, there are two fundamental departures from these equations. First, the interferogram is recorded up to a finite path difference $x = L$, not to infinity. According to section 5.2, this should limit the resolving power to about L/λ, or σL. We need to look at the instrument function before we can define $\mathscr{R}$ more exactly. Secondly, the Fourier transform is computed digitally, using discrete sampling points at intervals Δx in the interferogram to give $B(\sigma)$ at discrete spectral intervals. The relations in equation 7.21 should therefore be Fourier sums rather than integrals, with the mathematical consequence that the computed spectrum is replicated at wavenumber intervals determined by $1/\Delta x$.

7.7.1 INSTRUMENT FUNCTION AND RESOLUTION

We can find the instrument function by following the fortunes of a monochromatic input, $B(\sigma) \equiv B(\sigma_0)$. The interferogram is then

$$I(x) = B(\sigma_0) \cos 2\pi\sigma_0 x$$

The spectrum obtained from the transform, $S(\sigma)$, is

$$S(\sigma) = B(\sigma_0) \int_{-L}^{L} \cos 2\pi\sigma_0 x \cos 2\pi\sigma x \, dx$$

where we have assumed the interferogram to be recorded from $-L$ to $+L$ and have ignored for the moment the discrete sampling. The integrand is an even function of x, so

$$\begin{aligned} S(\sigma) &= 2B(\sigma_0) \int_0^L \cos 2\pi\sigma_0 x \cos 2\pi\sigma x \, dx \\ &= B(\sigma_0) \left[\int_0^L \cos 2\pi(\sigma_0 + \sigma) x \, dx + \int_0^L \cos 2\pi(\sigma_0 - \sigma) x \, dx \right] \\ &= B(\sigma_0) L \left[\frac{\sin 2\pi(\sigma_0 + \sigma) L}{2\pi(\sigma_0 + \sigma) L} + \frac{\sin 2\pi(\sigma_0 - \sigma) L}{2\pi(\sigma_0 - \sigma)} \right] \end{aligned} \tag{7.23}$$

Of these two terms, the first is negligibly small for all positive values of σ. We shall have to take account of it when we come to consider aliasing, but for the moment we ignore it and take the second term to represent the input line at σ_0. The L outside the brackets is a normalizing factor, and the instrument function is therefore the sinc function

$$\operatorname{sinc} 2\pi\sigma L \equiv \frac{\sin 2\pi\sigma L}{2\pi\sigma L} \tag{7.24}$$

This function is plotted in Fig. 7.13. At first sight the negative lobes are disconcerting, but of course the instrument function is a mathematical distortion of the original delta function input, *not* a real signal obtained by scanning a detector across the line. The integrated area (with the normalizing factor L) is equal to the height of the original delta function. The first zeros of the sinc function are at $\pm 1/2L$, and this distance is normally taken to define the resolution limit of the spectrometer:

$$\delta\sigma = 1/2L \Rightarrow \mathscr{R} = 2L/\lambda \tag{7.25}$$

This seemingly arbitrary choice is justified at the end of the section. As is evident from the derivation of equation 7.23, $\delta\sigma$ depends only on the extreme path difference L, not on the total length of the interferogram. The function is unchanged, except for a multiplicative factor $\frac{1}{2}$, if we take the integration limits from 0 to L instead of $-L$ to $+L$. $\delta\sigma$ is a somewhat optimistic estimate of the true resolution, for the FWHM of the sinc function, as measured from the zero line, is $1.2\delta\sigma$, and its side lobes, the first of which reach 20% of the peak, are undesirable features. This 'ringing' is a mathematical consequence of the abrupt truncation at $\pm L$ and can actually be modified. Indeed, an important feature of

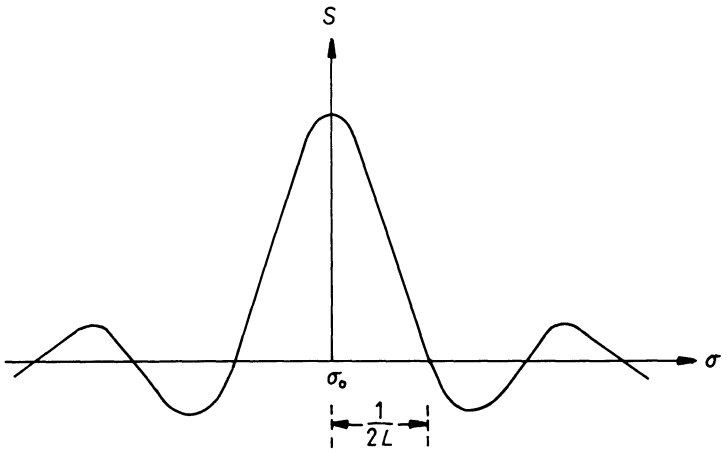

Figure 7.13 The intensity distribution $S(\sigma) = \operatorname{sinc} 2\pi(\sigma - \sigma_0) L$ for a monochromatic line at σ_0. The first zeros are at $\pm 1/2L$.

Fourier transform spectroscopy is the ability to vary the instrument function mathematically, as we shall now show.

One of the properties of Fourier transforms is that the transform of the product of two functions is the convolution of their individual transforms [7, 8, 10]. Let $A(x)$ represent the function which truncates the interferogram, so that the finite interferogram is described by $A(x)\,I(x)$. Then the recovered spectrum is given by

$$S(\sigma)=\widetilde{A(x)\,I(x)}=\tilde{A}(\sigma)*B(\sigma) \tag{7.26}$$

where the wavy line indicates the Fourier transform and $*$ means convolution. The instrument function $\tilde{A}(\sigma)$ is the Fourier transform of the truncation function $A(x)$. There is nothing peculiar to Fourier transform spectroscopy about the right-hand side of this equation: the output from a grating spectrometer or a Fabry–Perot can similarly be expressed as the convolution of the spectrum with the appropriate instrument function, the sinc^2 or the Airy distribution, respectively. The sinc function of equation 7.24 is the Fourier transform of the top-hat truncation function

$$A(x)=1 \quad \text{for } |x|\leqslant L \qquad A(x)=0 \quad \text{for } |x|>L$$

i.e.

$$\tilde{\sqcap}_{\mathrm{L}}(x)=L\,\text{sinc}\,2\pi\sigma L$$

Since the Fourier transform of $\cos(2\pi\sigma_0 x)$ is a delta-function pair at σ_0 and $-\sigma_0$, equation 7.26 for this case becomes

$$S(\sigma)=\widetilde{A(x)\,I(x)}=\text{sinc}\,2\pi\sigma L*[\delta(\sigma\pm\sigma_0)]$$

as shown schematically in Fig. 7.14. The mirror image of the spectrum at $\sigma=-\sigma_0$ corresponds to the first term in equation 7.23. It is not peculiar to this mythical monochromatic source function; the Fourier transform process always gives a mirror image $B(-\sigma)$ such that

$$B(-\sigma)=B(\sigma)$$

The instrument function can be varied by changing $A(x)$, which can be done mathematically, simply by multiplying each interferogram point by an appropriate factor. For example, if $A(x)$ is the triangular function Λ_{L} falling linearly from 1 at $x=0$ to 0 at $x=\pm L$, then $\tilde{A}(\sigma)=\text{sinc}^2\pi\sigma L$, which is identical to the diffraction-limited instrument function of a grating spectrometer. The distance to the first zero in this case is $1/L$, twice the value for the sinc function, and the FWHM is also increased, from $1.2\,(1/2L)$ to $\sim 1.8\,(1/2L)$, but the side lobes are reduced to about 4% of the maximum. For this reason the tapering of the truncation function to reduce the ringing is known as apodization – literally 'cutting off the feet'. Although the triangular function is the best-known example, presumably because it *does* reproduce the grating function, other truncation functions are nearly always preferable and can be chosen to suit particular problems. For a given value of L, the more gentle the cut-off the more the

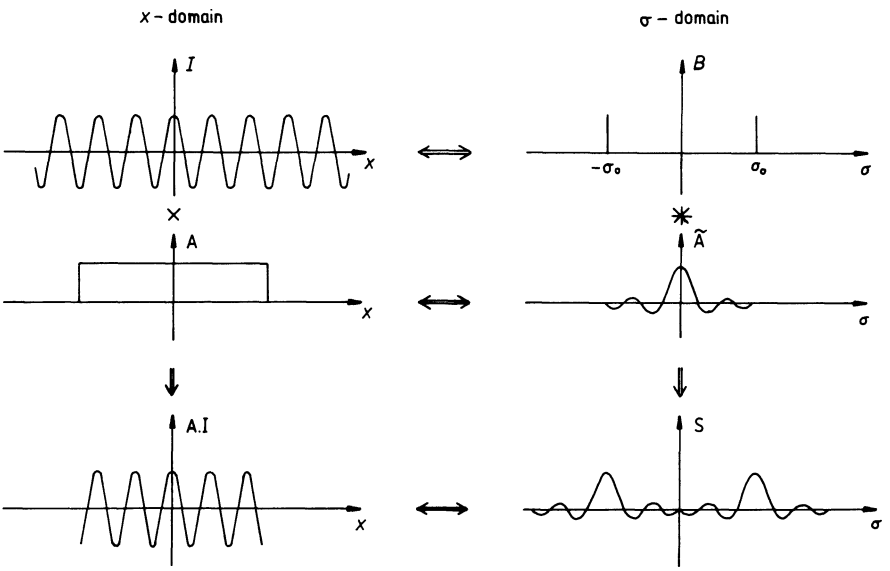

Figure 7.14 Multiplication of the infinite cosine interferogram by the top-hat truncation function $A(x)$ in the x-domain, corresponding to convolution of the delta-functions at $\pm\sigma_0$ with the sinc function $\tilde{A}(\sigma)$ in the σ-domain. The two functions in each horizontal row are Fourier transforms of one another.

ringing is reduced, though always at the expense of increased FWHM. If L is large enough for the true spectral linewidth to exceed $1/2L$, then the interferogram is effectively self-apodized – i.e. the modulation falls to zero at the ends of the interferogram, and the ringing disappears. In terms of equation 7.26, the convolution approximates to $B(\sigma)$ when $\tilde{A}(\sigma)$ is appreciably narrower than $B(\sigma)$. The line is then said to be fully resolved. Again, this situation is not peculiar to Fourier transform spectroscopy, but it is not so frequently encountered with gratings because of their lower potential resolving power.

7.7.2 SAMPLING AND ALIASING

We come now to the replication of the spectrum arising from the discrete sampling. This is a general feature of Fourier series. $B(\sigma)$ in equation 7.21 should actually be

$$B(\sigma) = \sum_{j=-N}^{N} I(j\Delta x)\cos 2\pi\sigma(j\Delta x) \tag{7.27}$$

Δx represents the 'fundamental frequency' in this Fourier series, and the 'pulse' $B(\sigma)$ is replicated at intervals $\Delta\sigma = 1/\Delta x$ in the wavenumber domain. This is

obvious if one writes out the integrand:

$$I(j\Delta x)\cos[2\pi j\Delta x(\sigma+\Delta\sigma)]=I(j\Delta x)\cos(2\pi\sigma j\Delta x+2\pi j)$$
$$=I(j\Delta x)\cos 2\pi\sigma j\Delta x$$

Therefore

$$B(\sigma)=B(\sigma+\Delta\sigma)=B(\sigma+2\Delta\sigma)=\cdots$$

The replication can also be understood from the convolution theorem, equation 7.26. Sampling the interferogram at intervals Δx means multiplying it by a so-called Dirac comb, or shah function, as shown in Fig. 7.15. The Fourier transform of this Dirac comb is another Dirac comb with spacing $\Delta\sigma=1/\Delta x$ [10]. This second comb has to be convoluted with the spectrum, producing the replication shown in Fig. 7.15.

We can now define a minimum 'safe' sampling interval. If the original spectrum extends from $\sigma=0$ to some maximum value, σ_m, the computed spectrum occupies a range $2\sigma_m$ because of the mirror image. As drawn in Fig. 7.15, the negative image of the second replication, marked 2^-, is just clear of the end of the piece we actually want, marked 1^+. It can be seen that this requires $\Delta\sigma=2\sigma_m$, or

$$\Delta x=\frac{1}{2\sigma_m}=\frac{\lambda_{min}}{2}, \tag{7.28}$$

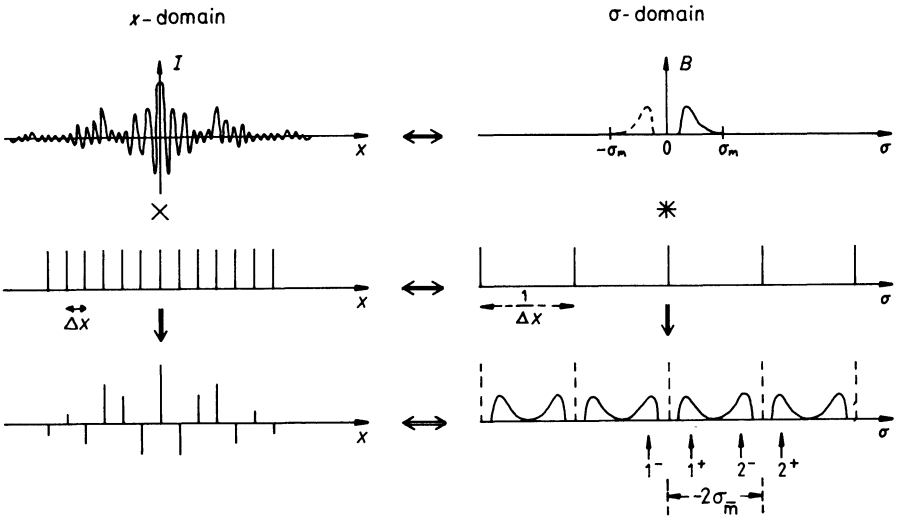

Figure 7.15 Sampling (i.e. multiplication) of the interferogram by the Dirac comb of spacing Δx in the x-domain, corresponding to convolution of the spectrum with the Dirac comb of spacing $\Delta\sigma=1/\Delta x$ (i.e. replication at intervals $\Delta\sigma$). As in Fig. 7.14 the two functions in each row are Fourier transforms of one another.

In other words, it is necessary to sample the interferogram at least every half-wavelength of the shortest wavelength in it to avoid some sort of overlap problem. This rule is an application of the more general *Nyquist sampling theorem.* If it is not satisfied the spectrum cannot be recovered unambiguously from the interferogram. For example, suppose the real spectrum does not stop at the limit σ_m in Fig. 7.15, but contains a feature at $1.5\,\sigma_m$. There is no way of distinguishing in the computed spectrum between this real line and the spectral region around $0.5\sigma_m$ corresponding to the inverted replication of 1^+ in the section 2^-.

This replication ambiguity is known as *aliasing.* It is a little reminiscent of order overlap in gratings, and to that extent $1/2\Delta x$ may be considered as the free spectral range; but it is a distinctive type of order overlap in which alternate orders are reversed. A convenient way of looking at it is shown in Fig. 7.16, where the aliases, each of width $\sigma_w = 1/2\Delta x$, shown strung out in Fig. 7.16(a), are folded back zig-zag in Fig. 7.16(b). The wavenumber of a feature appearing at X, corresponding to say $0.8\sigma_w$ in the first fold, might equally well be $1.2\sigma_w$, $2.8\sigma_w$, $3.2\sigma_w$, . . . unless it is known that the spectrum stops at σ_w. This diagram also shows how a spectrum can be deliberately aliased. If the source or detector characteristics or a suitable optical filter ensure that the detected radiation is all between, say, $2\sigma_w$ and $3\sigma_w$, then no ambiguity results from sampling at $\Delta x = \frac{1}{2}\sigma_w$ instead of the Nyquist value $\frac{1}{6}\sigma_w$, a saving of a factor of 3 in the number of points to be transformed.

Finally, we return to the question of the fundamental resolution limit. Suppose we were to invert the actual procedure by sampling the spectrum in order to compute the interferogram. For a spectral sampling interval of $\delta_1\sigma$ the computed interferogram is replicated at intervals $1/\delta_1\sigma$ in the x-domain, and this basic interval needs to be long enough to accommodate an interferogram of

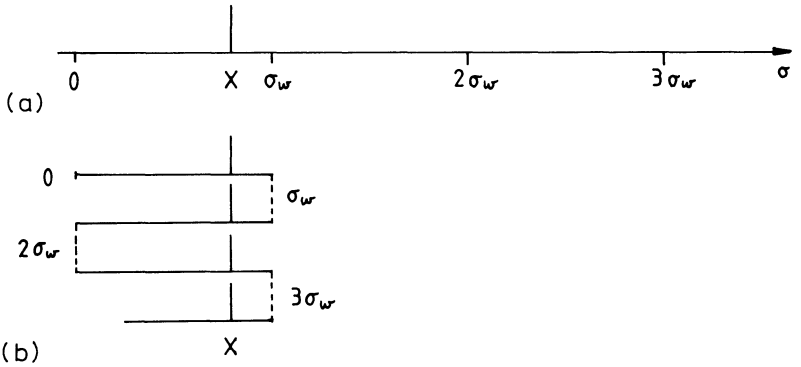

Figure 7.16 Representation of aliasing. The spectrum in (a) is folded back zig-zag by the sampling process in the manner shown in (b). The region σ_ω to $2\sigma_\omega$ corresponds to 2^- in Fig. 7.15, $2\sigma_\omega$ to $3\sigma_\omega$ corresponds to 2^+, etc. The feature X at $0.8\sigma_\omega$ might also be at $1.2\,\sigma_\omega$ in the second alias, $2.8\sigma_\omega$ in the third alias, etc.

length $2L$ ($-L<x<+L$). This requires $1/\delta_1\sigma = 2L$, or

$$\delta_1\sigma = 1/2L \Rightarrow \delta_1\sigma = \delta\sigma = 1/2L$$

in accordance with equation 7.25. If this mathematical argument seems physically artificial, the same result can be obtained by requiring a one-to-one correspondence between the number of independent input (interferogram) points and independent output (spectrum) points in the Fourier transform. For the input we take $N = L/\Delta x$, because the part of the interferogram, if any, between $-L$ and 0 actually duplicates the information between 0 and $+L$ (though the next section shows that there are good reasons in certain cases for taking it). For the spectrum the number of independent points is the number of resolution elements in the alias width, $N = \sigma_w/\delta\sigma$. Equating these two values of N gives

$$N = \frac{L}{\Delta x} = \frac{\sigma_w}{\delta\sigma} \tag{7.29}$$

so

$$\delta\sigma = \frac{\sigma_w \Delta x}{L} = \frac{1}{2L} \quad \text{by equation 7.28}$$

7.8 Fourier transform spectroscopy: computing and control

The increasing use of Fourier transform spectroscopy has been closely coupled with the development of digital computers. For a really simple spectrum, consisting of a few lines only, it is possible to deduce line separations and shapes from the visibility of the fringes in the interferogram, a method used successfully by Michelson and requiring considerable intuition. The advent of powerful and relatively cheap computers has rendered the visibility method obsolete.

All the same, the computing power required for a large Fourier transform should not be underestimated. For an N-point transform each spectral point requires N multiplications and N additions, so for N spectral points there are $2N^2$ operations for a full Fourier transform. This number is greatly reduced by the fast Fourier transform (FFT), a method which uses the symmetry properties due to the equal steps in both the interferogram and the spectrum to achieve the transform in $2N \log_2 N$ operations [11]. For $N = 1000$, this saves a factor of about 100 (since $\log_2 N \simeq 10$), and for $N = 10^6$ the saving is a factor of $\simeq 50\,000$. Even so, it is no trivial matter to store 10^6 data points, perform 40 million operations on them, and re-store them as spectral points. These very large values of N are a necessary consequence of high resolution. From equation 7.29, $N = \sigma_w/\delta\sigma$ – i.e. the number of spectral points into which the alias is divided; in the first alias N is numerically equal to the resolving power.

The demands imposed by large N are one of several reasons why Fourier transform spectroscopy has not in the past been considered suitable for the ultra-violet. For many purposes – finding energy levels of atoms or molecules,

for example – ultimate accuracy depends on the resolution limit $\delta\sigma$, rather than on $\mathscr{R}$, and, for constant $\delta\sigma$, N scales with σ_w, the largest wavenumber in the spectrum. Consequently, larger values of N are usually demanded in the visible and ultra-violet than in the infra-red.

So far we have implicitly assumed a symmetric interferogram, from which the spectrum can be accurately recovered by the cosine transform. In practice some degree of asymmetry is almost inevitable – for example, from a slight mismatch of the paths of the two beams through the beamsplitter substrate. An asymmetric interferogram, such as that shown in Fig. 7.11, can always be represented by the sum of two interferograms, one symmetric and the other antisymmetric. Only the former is recovered by the cosine transform, because from symmetry considerations the cosine transform of an odd function is necessarily zero. To recover the true spectrum it is necessary to take the sine transform as well. This is done by taking the complex Fourier transform. The pair of equations 7.21 then becomes

$$\left.\begin{aligned} I(x) &= \int_{-\infty}^{\infty} B(\sigma)\cos[2\pi\sigma x + \varphi(\sigma)]\,\mathrm{d}\sigma \\ S(\sigma) &= \int_{-\infty}^{\infty} I(x)\mathrm{e}^{-2\pi i\sigma x}\,\mathrm{d}x \end{aligned}\right\} \tag{7.30}$$

where $\varphi(\sigma)$ is the phase difference between the two beams due to the wavenumber-dependent part of the path difference. The inverse transform now yields the complex quantity

$$S(\sigma) = B(\sigma)\mathrm{e}^{-i\varphi(\sigma)}$$

To recover $B(\sigma)$ from $S(\sigma)$ one can either take the modulus or one can calculate $\varphi(\sigma)$ from the ratio of the imaginary and real parts and multiply each spectral point by $\mathrm{e}^{i\varphi(\sigma)}$ – i.e.

$$\left.\begin{aligned} &B(\sigma) = \{\mathrm{Re}[S(\sigma)]^2 + \mathrm{Im}[S(\sigma)]^2\}^{1/2} \\ &\text{or} \quad \varphi = \arctan\left(\frac{\mathrm{Im}[S(\sigma)]}{\mathrm{Re}[S(\sigma)]}\right) \quad \text{and} \quad B(\sigma) = S(\sigma)\mathrm{e}^{i\varphi(\sigma)} \end{aligned}\right\} \tag{7.32}$$

The calculation of $\varphi(\sigma)$ requires a symmetrically truncated part of the interferogram (since otherwise the instrument function $\tilde{A}(\sigma)$ itself contains a phase factor) so it is always necessary to record a section of the interferogram for $x < 0$. The demands on the accuracy of the phase correction are reduced if the interferogram can be recorded from $-L$ to $+L$. In certain applications it is actually the phase information that is required – for example, to measure the wavelength dependence of the refractive index of a specimen inserted in one of the two beams. In this technique, known as dispersive Fourier transform spectroscopy, it is the asymmetry of the interferogram that gives the desired information.

The FFT algorithm relies on the equality of the steps Δx in the interferogram samples. This is normally achieved by monitoring the path difference with a helium–neon laser and using the laser fringes either to specify the path length or to servo the drive velocity so that the samples can be taken at equal time intervals. For the visible and ultra-violet regions, where the wavelength is shorter than that of the laser, it is necessary to subdivide the laser fringes to satisfy the Nyquist sampling condition – in other words, to obtain unaliased spectra. Random errors in sampling produce 'white' noise; periodic errors give rise to ghosts with the same general characteristics as grating ghosts: the slower the period, the closer the ghost to its parent.

7.9 Fourier transform versus grating spectroscopy: a comparison

The advantages offered by Fourier transform spectroscopy (FTS) in the infra-red are significant enough to have made it an important technique even before computers were so widely available and so relatively cheap. There are also substantial advantages in the visible and near ultra-violet that are not so widely appreciated. The object of this section is to discuss the relative merits of the two types of instrument for various applications. Fabry–Perot interferometers have not been included in this comparison because they really serve rather a different purpose: that of achieving very high resolution over a limited wavelength range, and usually with the aid of a prism or low-resolution grating spectrograph.

In the far ultra-violet the grating has no rival. The Michelson interferometer is limited by the transmission of the beamsplitter plates to wavelengths greater than 170 nm – or possibly 110 nm with suitable crystal beamsplitters. All-reflection interferometers are possible in principle, but have not yet been realized in practice.

From 170 nm through the visible and the entire infra-red region the principal advantages of Fourier transform spectroscopy are higher potential resolving power and better signal-to-noise ratio for any given resolving power. The first comes about simply because the scan length L of an interferometer can always be made to exceed the width W of a grating (sections 7.7 and 6.3). There are two contributory factors to the second: the higher throughput and – in the infra-red only – the multiplex factor; these are often called, after the spectroscopists who first pointed them out, the Jacquinot and Fellgett advantages, respectively. We consider first the throughput advantage, which is common to all wavelength regions.

As shown in section 7.2, the maximum angular aperture that can be subtended by the circular input aperture of an interferometer for a given resolving power $\mathscr{R}$ is given by

$$\theta_m = \frac{a}{f} = \sqrt{\frac{2}{\mathscr{R}}}$$

(equation 7.4). This limit is frequently expressed in terms of the maximum solid angle Ω_m that can be subtended by the aperture at the collimating lens:

$$\Omega_m = \frac{\pi a^2}{f^2} \Rightarrow \mathscr{R}\Omega_m = 2\pi \tag{7.33}$$

In terms of the throughput defined in equation 4.5 (input aperture area × input solid angle) we have

$$T_i = \pi a^2 \frac{A_i}{f^2} = \Omega_m A_i$$

where A_i is the area of the collimating lens or the beamsplitter, so

$$\mathscr{R}T_i = 2\pi A_i \tag{7.34}$$

For the slit-limited grating throughput, from equation 6.14:

$$\mathscr{R}T_g \simeq \beta WH = \beta A_g \tag{7.35}$$

where β is the angular height of the slit and A_g the area of the grating. The throughput advantage is often expressed by assuming equal areas A for the two central components (the beamsplitter and the grating), so that

$$\frac{T_i}{T_g} = \frac{2\pi}{\beta} \gtrsim 100 \tag{7.36}$$

since β seldom exceeds 0.01 (section 6.3). This comparison is superficially a little unfair because gratings are usually larger than beamsplitters, but essentially the same result is obtained by comparing the two instruments both working at the same input f number – i.e. the same input solid angle. In this case

$$\frac{T_i}{T_g} = \frac{A_i'}{A_g'} \quad \text{where} \quad \begin{cases} A_i' = \pi a^2 = \dfrac{2\pi f^2}{\mathscr{R}} & \text{(equation 7.4)} \\ A_g' = wh = \dfrac{2f \tan\theta\, h}{\mathscr{R}} & \text{(equation 6.13)} \end{cases}$$

so for $\tan\theta \simeq 1$ and the same $\mathscr{R}$

$$\frac{T_i}{T_g} \simeq \frac{\pi f}{h} = \frac{\pi}{\beta} \tag{7.37}$$

The multiplex advantage applies only when the noise is predominantly detector noise, a situation that used to prevail throughout the infra-red but is no longer necessarily true in the near infra-red (sections 4.10 and 4.11). If a noisy detector is used with a grating to scan sequentially through N spectral elements in a total time T, the integration time per element is T/N. A Fourier transform spectrometer observes all these spectral elements simultaneously for the full scan time T for no additional noise penalty, thereby increasing the integration time by a factor N and the signal-to-noise ratio by a factor $\sqrt{N}$.

The relative merits of the two instruments when the noise is not detector-limited are a little less well defined. Photon noise is equal to the square root of the signal, so in photon noise-limited conditions, broadly speaking, the increase of signal by a factor N at each scanning point from the multiplexing increases the noise by a factor $\sqrt{N}$, thereby cancelling the factor gained from integration time. However, the noise distribution in the two spectra is sufficiently different to require further scrutiny. Let Φ_σ be the detected photon flux (in photons s^{-1}) in any given spectral element at wavenumber σ. For a total observation time T, a scanning grating spends time T/N on each element and accumulates $\Phi_\sigma T/N$ photons. The signal-to-noise ratio is thus

$$\frac{s_\sigma}{n} = \sqrt{\frac{\Phi_\sigma T}{N}} \tag{7.38}$$

The interferometer also spends time T/N at each point of the scan, but collects each time $\sum\Phi_\sigma$ photons s^{-1}. The signal-to-noise in the interferogram is therefore

$$\left(\frac{\bar{s}}{n}\right)_x = \sqrt{\frac{\sum\Phi_\sigma T}{N}} = \sqrt{\bar{\Phi}T} \tag{7.39}$$

where $\bar{s}$ and $\bar{\Phi}$ are the mean signal and photon flux, respectively. Now the mean signal is not changed by the transform – there are still on average $\bar{\Phi}$ photons s^{-1} per spectral point – but if the noise is 'white' each point in the interferogram contributes noise to each point in the spectrum. The contributions add incoherently to increase the noise by a factor $\sqrt{N}$, giving

$$\left(\frac{\bar{s}}{n}\right)_\sigma = \frac{1}{\sqrt{N}}\left(\frac{\bar{s}}{n}\right)_x = \sqrt{\frac{\bar{\Phi}T}{N}} \tag{7.40}$$

In contrast with the grating, the *noise is the same in all spectral elements*, regardless of whether they have signal in them. The signal-to-noise for a particular element is found by multiplying by $s_\sigma/\bar{s}$, giving finally

$$\frac{s_\sigma}{n} = \frac{s_\sigma}{\bar{s}}\left(\frac{\bar{s}}{n}\right)_\sigma = \frac{s_\sigma}{\bar{s}}\sqrt{\frac{\bar{\Phi}T}{N}} \tag{7.41}$$

For a quasi-continuum, where $\Phi_\sigma \simeq \bar{\Phi}$ and $s_\sigma \simeq \bar{s}$ everywhere, equation 7.41 is identical with equation 7.38 for the grating. For a sparse emission spectrum $s_\sigma/\bar{s}$ can be very large at the line positions, giving an apparent advantage to the FTS – but, of course, a grating spectroscopist is unlikely to spend much time scanning empty spectral regions. Generalizations about the multiplex advantage or disadvantage in the photon-limited case must be treated with caution.

If the noise is dominated by source fluctuations, there really is a multiplex disadvantage. Suppose the noise is some constant fraction ε of the signal. The signal-to-noise for both grating and interferometer is $1/\varepsilon$, and by virtue of

equations 7.40 and 7.41 we have for the FTS

$$\frac{s_\sigma}{n}=\frac{s_\sigma}{\bar{s}}\frac{1}{\sqrt{N}}\frac{1}{\varepsilon}$$

which is on average worse by a factor of $1/\sqrt{N}$ than for the grating.

Finally, there are considerations other than signal-to-noise ratio. A high-resolution FTS is a more compact instrument than its grating counterpart, and it has an instrument function that is wavelength-independent and can be modified at will, plus a built-in wavelength scale from the sampling control laser. On the other hand, it requires a computer to interpret the spectrum, the optical and mechanical tolerances at short wavelengths are very stringent, and it cannot be used as a monochromator. Some problems can be tackled only by one or other type of instrument, while for others the advantages may be fairly evenly balanced.

References

INTERFEROMETRIC SPECTROSCOPY IN GENERAL

1. Bousquet, P. (1971) *Spectroscopy and its Instrumentation*, Hilger, London.
*2. Francon, M. (1966) *Interferometry*, Academic Press, New York.
*3. Steel, W.H. (1967) *Interferometry*, Cambridge University Press (2nd edn 1983).
*4. Tolansky, S. (1973) *Introduction to Interferometry*, (2nd edn), Longmans, London.

FABRY–PEROT INTERFEROMETERS

5. Hernandez, G. (1986) *Fabry–Perot Interferometers*, Cambridge University Press.
6. Tolansky, S. (1947) *High Resolution Spectroscopy*, Methuen, London.

FOURIER TRANSFORM SPECTROSCOPY

7. Bell, R.J. (1972) *Introductory Fourier Transform Spectroscopy*, Academic Press, New York.
8. Chamberlain, J.C. (1979) *The Principles of Interferometric Spectroscopy*, Wiley, New York.
9. Articles in *Spectrometric Techniques* (ed. G.A. Vanasse), Academic Press, New York:
   Bell, R.J. (1977) Applications of Fourier transform spectroscopy, Vol. I.
   Breckinridge, J.B. and Schindler, R.A. (1981) First-order optical design for Fourier spectrometers, Vol. II.
   Sakai, H. (1977) High resolving power Fourier spectroscopy, Vol. I.

MATHEMATICAL BACKGROUND TO FOURIER TRANSFORMS

10. Bracewell, R. (1965) *The Fourier Transform and its Applications*, McGraw-Hill, New York.
11. Brigham, E.O. (1974) *The Fast Fourier Transform*, Prentice-Hall, Englewood Cliffs, New Jersey.

* Further reading, not cited in text.

EIGHT

Laser spectroscopy

The principal difficulty associated with a chapter on laser spectroscopy is what to leave out. Many entire books are devoted to techniques and applications of lasers in atomic and optical physics. We have tried to stick to a guideline of direct relevance to optical spectroscopy, particularly high-resolution spectroscopy, and to give an idea of the types of experiment that can be performed 'better' with lasers and the types that cannot be performed without lasers. Certain applications that fall in the first category, double resonance and Raman spectroscopy and the measurement of lifetimes of excited states, have been left to the appropriate sections of Chapters 9 and 12. No attempt is made to cover non-linear effects like self-induced transparency, optical bistability and the optical transients such as superradiance, optical nutation and photon echoes.

Bear in mind that some of what is written about such a rapidly developing field must inevitably be out of date by the time it is read, and other important techniques will have been initiated and developed in the meantime. The same problem of rapid development applies to the reference list. Books on laser spectroscopy tend to be collections of 'state of the art' papers, and good review articles are few. We have listed all of the latter that we could find, and the former should at least guide the reader to sources of up-to-date information.

8.1 Introduction to lasers [1, 2]

The action of masers and lasers depends on stimulated emission, as represented by the Einstein coefficient B_{21} introduced in section 4.5. Maser stands for *microwave* amplification by stimulated emission of radiation and laser for *light* ditto. Consider the 'two-level atom' illustrated in Fig. 4.4. In the presence of radiation of density $\rho(\nu_{12})$, every atom in level 1 has a probability per unit time $B_{12}\rho$ of capturing a photon and attenuating the radiation, whereas every atom in level 2 has a probability $B_{21}\rho$ of adding a photon and amplifying the radiation. The added photon has exactly the same energy and direction as the stimulating photon; in terms of electromagnetic theory, the waves add coherently, with the same wave vector and phase.

Normally there are more atoms in level 1 than in level 2 – by a factor $(g_1/g_2)\exp(h\nu_{12}/kT)$ in thermodynamic equilibrium (equation 4.17) – and there is net absorption of radiation. The condition for amplification is

$$N_2B_{21}\rho > N_1B_{12}\rho$$

Using equation 4.19 to relate the B-coefficients, this becomes

$$N_2/N_1 > g_2/g_1$$

This condition, known as population inversion, clearly cannot be realized in thermodynamic equilibrium, for which the best that can be achieved as $T \to \infty$ is the statistical balance g_2/g_1:

$$\frac{N_2}{N_1} = \frac{g_2}{g_1} e^{-h\nu_{12}/kT} \to \frac{g_2}{g_1} \quad \text{as } T \to \infty \tag{8.1}$$

Moreover, population inversion is a necessary but not a sufficient condition for amplification because atoms that decay spontaneously (incoherently) are lost from the coherent amplification process. In the remainder of this section we discuss briefly both how population inversion can be achieved and how the spontaneous emission loss can be offset by enclosing the radiation in a resonant cavity.

It is impossible by any process to invert the population in an isolated two-level atom: the best that can be done by radiative excitation is the infinite temperature (or infinite ρ) state of equation 8.1, and it is shown in Chapter 13 that the same is true of collisional excitation. It is therefore necessary to feed level 2 from some other level 3. Figure 8.1 shows the mechanism for two

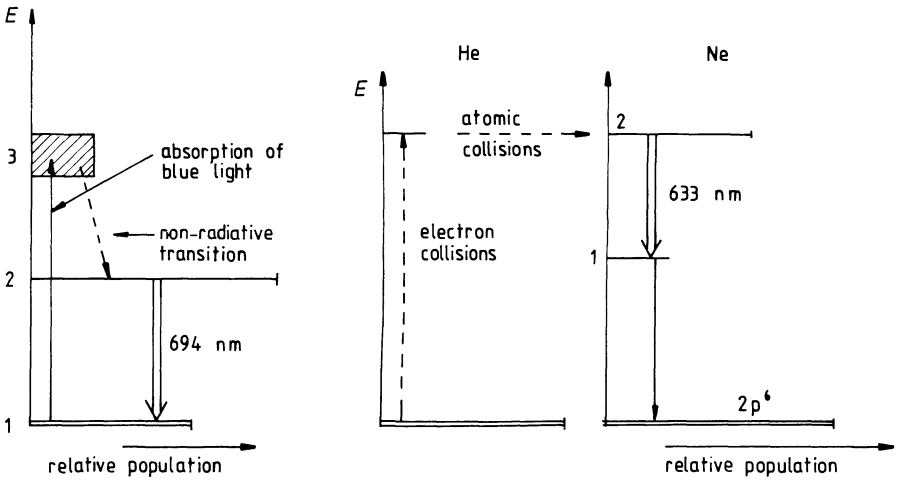

Figure 8.1 Population mechanisms for (a) ruby and (b) helium–neon laser. The double arrow shows the lasing transition in each case.

common types of laser, ruby and helium–neon. In ruby the characteristic red radiation at 694 nm, like the red colour of the gemstone, comes from the small fraction of chromium ions (Cr^{3+}) present as an impurity in a sapphire (Al_2O_3) crystal. Energy levels in a solid are usually broadened by interactions with the neighbouring atoms or ions, but the lasing transition involves two levels in the partly filled 3d-shell, which are relatively sharp because they are shielded from the crystal fields by the outer electrons of the ion. Level 3, however, is broad, and this can be populated by a broad band of blue/near ultra-violet radiation from a flash discharge. Level 3 decays very rapidly by a radiationless process, giving up its energy to the lattice, to level 2, which is metastable. The population of level 2 piles up until the radiation density is sufficient to trigger a 'chain reaction' of stimulated emission, which continues until levels 1 and 2 reach the statistical balance condition of equation 8.1. The laser necessarily operates in pulses because of the excitation mechanism and the need for dissipation of lattice energy. The He–Ne laser, on the other hand, is a continuously running (usually known as continuous wave, or CW) laser. Broadband excitation for a gas laser, where all the levels are sharp, is necessarily very inefficient. In this case level 2 of the lasing transition in Ne is populated by collisions with He metastable atoms of the same energy, excited by electron collisions in a continuously running discharge. The coincidence of the relevant energy levels in the two atoms makes for very efficient energy transfer, the He atom returning to the ground state after collision. Since the $2 \rightarrow 1$ laser transition ends on an excited state which decays rapidly to lower states, the population of 1 does not build up, and population inversion can be maintained without 2 being metastable. Actually both 1 and 2 comprise more than one level, and the discharge can be made to lase on several infra-red lines in addition to the well-known red line at 633 nm.

Losses from spontaneous emission increase with ν^3 (equation 4.19), and resonant cavities, which are not necessary for masers or for some far infra-red lasers, are essential in the optical region. The cavity is simply a pair of parallel mirrors, one at each end of the column of material in which amplification is to take place. One mirror has 100% reflectivity, while the other transmits a certain fraction – say 10–50%, depending on the type of laser – to form the output. The length of the cavity is tuned to an exact number of half-wavelengths of the lasing transition so that standing waves are formed, and the radiation density is higher than that attainable from a single passage of a travelling wave. The geometry of the cavity ensures that the build-up of radiation applies only to a small solid angle centred on the cavity axis. Ideally the emerging beam has a divergence limited by the diffraction spread λ/D, where D is its diameter. This can be as little as 10^{-4} rad, or 10 cm in 1 km, but divergences of the order of a milliradian are more common. The spectral bandwidth of the radiation is determined by the amplification of the cavity – i.e. its Q-value. The radiation from an excited atom is normally spread over a range of frequencies corresponding to Doppler and perhaps pressure broadening (Chapter 10). However, only certain frequencies within this band are amplified by the resonant cavity, as shown in Fig. 8.2. The

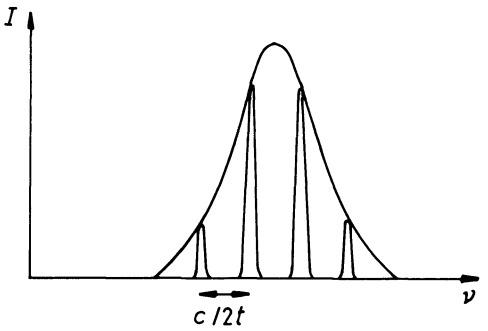

Figure 8.2 Laser gain curve. The broad curve is the gain curve of the lasing medium and the narrow curves are the longitudinal modes. For a cavity of length t the frequency separation of these is $c/2t$.

laser is said to be running multi-mode if two or more of these frequencies are contained in the output. By differentiating the constructive interference condition for a cavity of length t, $2t = m\lambda$, and putting $\Delta m = 1$, it can be seen that the mode spacing is given by

$$\delta\lambda = \frac{\lambda}{m} = \frac{\lambda^2}{2t} \quad \text{or} \quad \delta\nu = \frac{c}{2t} \tag{8.2}$$

For a sufficiently short cavity only one mode is contained in the envelope of Fig. 8.2, and the laser runs mono-mode. The wavelength spread is then limited essentially by the thermal noise or stability of the cavity. The various ways of 'locking' to a fixed frequency, are discussed in section 8.4.

Although lasers of fixed frequency have found a few applications in spectroscopy – notably in Raman spectroscopy (sections 9.10 and 9.11) – it was the development of tunable lasers that really revolutionized the subject. The availability of high power at low divergence, together with narrow spectral bandwidth or very short pulse times, at almost any desired wavelength, made it possible both to improve the accuracy of conventional measurements of such quantities as small wavelength differences, line profiles and lifetimes, and to devise new measurement techniques such as Doppler-free spectroscopy. In addition, a number of entirely new fields – multiphoton spectroscopy and interactions at high levels of radiative power, for example – have been opened up and explored.

8.2 Types of laser [1–4, 8–10, 30]

8.2.1 FIXED FREQUENCY

The direct applications of fixed-frequency lasers to spectroscopy are almost limited to Raman-related techniques (Chapter 9), but they have a number of

auxiliary uses. The familiar helium–neon laser described above is an invaluable aid to alignment and adjustment of optical instruments. High-power pulsed lasers are used to generate blobs of plasma as sources of high stages of ionization, as mentioned in section 4.8.4. Of more relevance to this chapter is the use of fixed frequency lasers to pump tunable dye lasers operating in the visible or near ultra-violet. Lasers operating at wavelengths much above 1 μm are not suitable for this purpose, and we shall therefore ignore the many gas lasers that operate on vibration–rotation transitions in the infra-red – for example, the well-known CO_2 laser transition at 10.6 μm.

The important gas lasers in the visible and ultra-violet regions are the nitrogen laser (pulsed) and the continuously running (CW) inert gas ion lasers. N_2 lases at 337 nm on an electronic transition, the upper level of which is populated by a capacitative discharge. The laser can deliver a power of about 1 MW over a few ns, by which time the population of the lower level has built up to statistical equilibrium; the molecules then decay back to the ground electronic state ready for the next pulse. Argon ion lasers operate CW on several wavelengths from the green to the near ultra-violet regions. The argon is ionized and excited by a high-density dc current, and lasing takes place between two sets of excited levels, the lower of which decays to the ground state rapidly enough to avoid population build-up. Krypton ion lasers operate in a similar fashion, but the corresponding transitions are at somewhat longer wavelengths, giving lines through the visible region. Output powers are a few watts.

The solid-state lasers that are important for our purpose are ruby, lasing on the Cr^{3+} transition at 694 nm as already described, and the rare earth neodymium. Transitions between levels of the partly filled inner shell of Nd^{3+} give rise to sharp lines in the same way as for Cr^{3+}. The ions may be incorporated either in a glass rod (Nd–glass) or in a crystal of yttrium–aluminium-garnet (Nd–YAG). The lasing mechanism is basically the same as that shown in Fig. 8.1 for ruby, except that the lower level of the lasing transition at 1.06 μm is itself an excited state, which decays rapidly to the ground state via a radiationless transition. Consequently, it is possible to run Nd–YAG lasers CW, although much higher powers can be obtained in the pulsed mode.

8.2.2 DYE LASERS

The tunable dye laser in one form or another is the nub of laser spectroscopy. Most dyes have absorption bands in the blue or ultra-violet and fluorescence bands at longer wavelengths. The energy level scheme is shown in Fig. 8.3. The molecules are pumped by radiation in the absorption band from the ground (singlet) state to a high vibrational level of another singlet state. They decay by radiationless transitions down the vibrational ladder until a radiative transition takes them down to one of the vibrational levels of the lower singlet state, from which they can trickle down to the ground state again. The path via radiationless transitions to and from the triplet state is a loss mechanism which can be

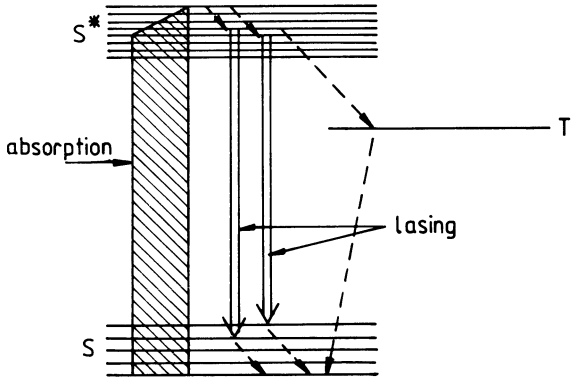

Figure 8.3 Energy levels of dye laser. The vibrational levels of the excited singlet state S^* are populated by broadband absorption from the levels of S, and lasing occurs over a band of vibrational levels represented by the two double arrows. Depopulation via the triplet levels T is a loss mechanism.

inhibited by appropriate solvents and additives. The dye can be pumped either by a flash discharge (usually a capacitative discharge through xenon) or by another laser. In CW use this is an Ar^+ or a Kr^+ laser, and in pulsed operation it may be a N_2 laser. Ruby and neodymium lasers can also be used to pump dyes, but their outputs have first to be frequency doubled or tripled (section 8.2.4) to bring them into the absorption band of the dye.

Dyes fluoresce over a wavelength band 20–50 nm wide. Narrow-band spectral output is achieved by tuning the resonant cavity containing the dye. In principle this can be done simply by replacing the fully reflecting mirror at one end by a diffraction grating, rotated to diffract the chosen wavelength back along the axis; in practice slightly more-complicated arrangements are adopted for reasons which we shall not go into here. It is also possible to tune by including one or more Fabry–Perot etalons of suitable spacing in the cavity. Because of the non-linear nature of the stimulated emission process, the radiation builds up rapidly at the particular wavelengths selected by the tuning element, so nearly all the energy from the untuned broad band can be channelled into a narrow spectral pass band. Typically, one might expect from a CW laser an output of a few hundred mW in a bandwidth of ~ 1 MHz, tunable over 30 GHz steps. For pulsed dye lasers the peak power can be a few hundred kilowatts with a pulse duration, depending on that of the pump laser, from a few ns to 1 μs.

8.2.3 TUNABLE LASERS IN THE INFRA-RED AND ULTRA-VIOLET

The range of tunable dye lasers can be extended to both longer and shorter wavelengths by frequency mixing, as described in section 8.2.4. Leaving aside

this technique for the moment, we find three types of tunable laser for the infra-red. Most recently developed are the colour centre or F-centre lasers, which are crystals of the heavier alkali halides doped with Li or Na (F_A and F_B centres, respectively). They are pumped and tuned in the same way as dye lasers, each type of crystal having an absorption band at shorter wavelength than its emission band. They cover the range from the far red to about 4 μm, and their tuning ranges and output powers are comparable with those of dye lasers. The other two types – semiconductor diodes and spin-flip Raman lasers – have been in use for longer. The diodes are p–n junctions, activated by a current across the junction. The resonant cavity is formed by the polished faces of the crystal, and is tuned by varying the temperature or applying a magnetic field. The spectral bandwidth of the output is very narrow, but the output power is only a few hundred μW. Spin-flip Raman lasers have output powers of the order of 1 W. The laser mechanism is Raman scattering (section 9.10) by electrons in a semiconductor in an external magnetic field. The energy difference between the incident and scattered radiation corresponds to alignment of the electron spins parallel or antiparallel to the applied field, and the frequency of the coherently scattered radiation can be tuned by varying the field. All three types of laser have to be cooled below room temperature.

For the ultra-violet we return to gas lasers in the form of excimers. The word is a portmanteau of 'excited dimer', the first such lasers being the inert gas dimers Xe_2, Kr_2 and Ar_2. The word is also applied, semantically incorrectly, to inert gas–halogen combinations such as KrF and XeF. The common characteristic of all these 'dimers' is a stable excited state and an unstable ground state. The dimers are formed in a highly excited state by an electron beam of high current density in gas at a pressure of a few atm, and they then cascade down to the lowest excited electronic state. For Xe_2, for example, this is Xe $(5p^5 6s)$ + Xe$(5p^6)$. There is always population inversion of this state with respect to the ground state because emission to the latter is followed immediately by dissociation. Emission takes place over a band of some tens of nm on the long-wavelength side of the atomic resonance line, as described in section 3.8.2 and illustrated in Fig. 3.12(c). The required spectral band is selected by a tuning element in the optical cavity. Excimer lasers operate in pulses of a few ns with a peak power that is usually of the order of 1 MW, depending on bandwidth, excimer type, etc. They are used both directly as tunable ultra-violet lasers, and also in frequency mixing and frequency doubling and tripling arrangements to reach wavelengths in the extreme ultra-violet. The inert gas monohalide lasers, which give very large powers at wavelengths around 200–300 nm, also serve as pumps for dyes with absorption bands in the ultra-violet.

8.2.4 FREQUENCY DOUBLING AND MIXING

Frequency mixing exploits the non-linear terms in the expression for the polarizability of a medium. The electric dipole moment $\boldsymbol{p}$ induced by an

electromagnetic wave, $\boldsymbol{E}=\boldsymbol{E}_0 \cos \omega t$, is to first order proportional to $\boldsymbol{E}_0$, and the polarizability and refractive index of the medium are usually calculated on the basis of this linear relationship. However, if $\boldsymbol{E}_0$ is large, higher-order terms must be included, and the expression for $\boldsymbol{p}$ becomes

$$\boldsymbol{p}=\alpha \boldsymbol{E}+\beta \boldsymbol{E}^2+\gamma \boldsymbol{E}^3+ \ldots \tag{8.3}$$

where α, β and γ are tensors unless the medium is isotropic. This expression appears again in section 9.11 in connection with CARS, and the generation of sum-and-difference frequencies is explained in a little more detail there. It can be seen that the second term in equation 8.3, $\beta \boldsymbol{E}_0^2 \cos^2 \omega t$, when rewritten as $\beta \boldsymbol{E}_0^2(\cos 2\omega t+1)$, contains the second harmonic of the original wave. In the same way 3ω appears in the third term. If all the dipoles in the medium are oscillating coherently, the emerging wave contains these harmonics along with the original frequency. For β to be non-zero the medium must have some anisotropy, a condition that can be met in crystals but not in gases. Moreover, for the coherence condition to be satisfied the refractive index of the medium has to be the same for the waves at ω and 2ω (or 3ω), a condition known as phase-matching.

Frequency doubling and tripling is often used to enable ruby or neodymium lasers to pump dyes, many of which have absorption bands corresponding to half the ruby or one-third the neodymium wavelengths. It is also used to extend the range of tunable dye lasers into the ultra-violet. Starting with an excimer laser and using suitable mixtures of metal vapour and inert gas for the tripling and phase-matching, it is possible to achieve tunable laser radiation even below 100 nm.

In frequency mixing the same process is applied to radiation from two lasers at different frequencies ω_1 and ω_2. The second-order term in equation 8.3 contains the sum-and-difference frequencies $\omega_1 \pm \omega_2$. If ω_1 represents a fixed frequency, the output can be tuned by varying ω_2 provided the phase-matching conditions for all three waves are met. Similarly, the third-order term contains not only the third harmonics, but also the combinations $2\omega_1 \pm \omega_2$ and $\omega_1 \pm 2\omega_2$. Evidently, frequency mixing can be used to generate tunable radiation in both the ultra-violet and the infra-red.

8.3 Optical pumping and saturation [1, 2, 16, 30]

Optical pumping is the name given to the process of increasing the population of an atomic or molecular level above its equilibrium value either by absorption of radiation or by a combination of absorption and re-emission. We have already met it in this chapter as one of the ways of achieving population inversion in a laser; more generally, it is a way of preparing a collection of atoms for further experiments or observations. In the double resonance and level-crossing experiments described in sections 9.8 and 9.9, for example, atoms are optically pumped

to a particular excited level, and a change in the characteristics of the re-emitted radiation is observed when they are transferred to another nearby level. Optical pumping experiments were carried out long before lasers existed, but they are much easier with tunable lasers because of the very high photon flux density within the absorption bandwidth of the transition that is being pumped.

As an example of the use of optical pumping to prepare a true two-level atom, consider the process illustrated in Fig. 8.4, which represents an atom with two states having $J=1$ and 2, respectively, in a magnetic field. There are three magnetic sub-levels in the lower state and five in the upper (section 2.9). If the atoms are irradiated with right-circularly polarized light of the correct frequency, they are excited to one of the upper sub-levels corresponding to $\Delta M_J = +1$. For re-emission $\Delta M_J = -1,\ 0,\ +1$ are equally probable, but the next absorption again increases M_J by 1. All the atoms are eventually pumped into the $M_J=1$ sub-level, and while the radiation is still on they continue to cycle between this level and the $M_J=2$ level of the upper state. A similar process can be applied to hyperfine magnetic sub-levels.

With high enough flux density at the resonance frequency it is possible to 'saturate' a transition. From Fig. 4.4 and equation 4.18 spontaneous emission can be neglected compared with stimulated emission if

$$B_{21}\rho \gg A_{21}$$

The equilibrium condition then becomes

$$B_{21}\rho N_2 = B_{12}\rho N_1$$

Using equation 4.19,

$$\frac{N_2}{N_1} = \frac{g_2}{g_1} \quad \text{for} \quad \rho \gg \frac{A_{21}}{B_{21}} = \frac{8\pi h\nu^3}{c^3}$$

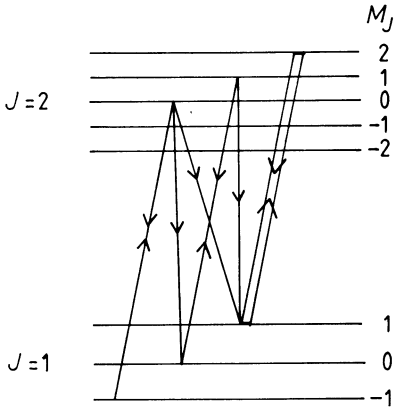

Figure 8.4 Example of optical pumping. Absorption of right-circularly polarized light corresponds to $\Delta M_J = +1$, and ideally all atoms are eventually cycled between the two levels joined by the double arrows.

In terms of $I(\nu_{12})$, the spectral flux density of a collimated beam (W m^{-2} Hz^{-1}), the saturation condition is

$$I \gg \frac{8\pi h\nu^3}{c^2} = \frac{8\pi}{\lambda^2} h\nu \tag{8.4}$$

which can be rather more vividly pictured as 8π photons per second per Hz over an area of one square wavelength. Once saturation is attained, the gas becomes transparent to radiation of that particular wavelength because absorption and stimulated emission are equally likely. Actually equation 8.4 is a necessary but not a sufficient condition for saturation; the laser has to supply sufficient power to replenish the losses from spontaneous emission, and in a gas of high column density this may be a more stringent criterion. If one assumes that all spontaneously emitted photons escape from the gas (i.e. radiation trapping is neglected), the power loss from a column of gas with area S (m^2), length L (m) and number density N (m^{-3}) is $NSLA_{21}$ photons s^{-1}. The laser spectral flux density required to keep pace with this is

$$I = \frac{NLA_{21}h\nu}{\Delta\nu} \quad \text{W m}^{-2}\,\text{Hz}^{-1}$$

where $\Delta\nu$ is the effective linewidth for absorption.

Saturation can be exploited in a number of ways, some of which appear later in this chapter. To understand them it is necessary to anticipate some of the discussion of linewidths in Chapter 10, in particular the sections on natural and Doppler broadening. In laser spectroscopy these are often referred to as homogeneous and inhomogeneous broadening, respectively. The reason for the distinction is that Doppler (inhomogeneous) broadening is the result of observing atoms with a range of line-of-sight velocities. If a particular subset of atoms all having essentially the same velocity can be selected in some way, the linewidth is determined by pressure effects or natural lifetime, and the line is said to be homogeneously broadened. The significance of the distinction for saturation is that narrow-band laser radiation at a frequency ν_1 travelling in the x-direction is absorbed only by the particular subset of atoms with an x-component of velocity v_1 such that ν_1 is Doppler-shifted to the central frequency ν_0 of the absorption line:

$$\frac{v_1}{c} = \frac{\nu_0 - \nu_1}{\nu_0} \tag{8.5}$$

It is thus possible to saturate a narrow band within the Doppler profile, the width of which is determined by whichever is the greater of the laser bandwidth and the homogeneous linewidth of the absorbing atoms. This process is known as 'hole-burning' and is illustrated in Fig. 8.5.

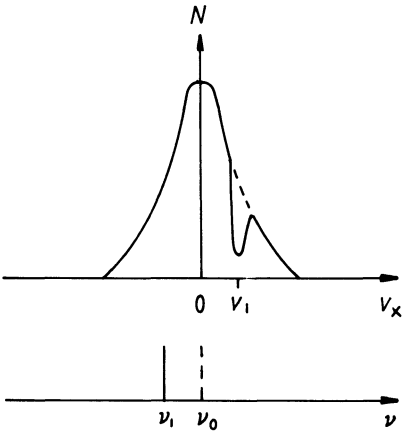

Figure 8.5 'Hole-burning'. Laser radiation at ν_1 is absorbed by atoms with a v_x-component of v_1, creating a 'hole' in the ground-state population at this value of v_1.

8.4 Spectral and temporal resolution [13, 15, 16, 29, 30]

The limiting accuracy with which one can measure line separations or positions or determine line profiles in a 'Doppler limited' experiment is governed by linewidths of order 1 GHz (0.03 cm^{-1}) in the visible region (section 10.3). Doppler broadening limits the useful resolving power of a spectrometer to a million or two, the value for which the instrumental width is smaller than the source linewidth. For the same reason it would be pointless to try for very narrow laser bandwidths were it not for the possibility of performing 'Doppler-free' measurements by one of the techniques described in section 8.6. These techniques can have as a limit the homogeneous linewidth of the transition, which may be two or three orders of magnitude smaller than the Doppler width. It is therefore useful to see how 'sub-Doppler' laser bandwidths can be realized.

There is a fundamental relation between spectral and temporal resolution set by the bandwidth theorem, $\Delta\nu\Delta t \sim 1$: a pulse of at least 10 ns is required for a linewidth of 100 MHz, for example. Instability and shot-to-shot reproducibility also affect the linewidths of pulsed lasers, and the narrowest laser linewidths are obtained with CW lasers.

A good CW dye laser can have an inherent linewidth of about 10 MHz, and since this 'width' is really due to small fluctuations, thermal and otherwise, in the length of the cavity it can be improved – to about 1 MHz – by locking the frequency to a stabilized reference cavity. One can think of this as a Fabry–Perot etalon of large separation, through which some fraction of the laser output is passed. If the laser frequency wanders, the transmission of this cavity changes, and a correction signal is generated to the laser cavity. The reference cavity itself may be locked to a second, fixed-frequency, stabilized laser. In normal use the

dye laser has to be scanned through the spectral region under investigation. This can be done either by changing the optical path in the reference cavity or by using this cavity as a Fabry–Perot to record successive maxima. The continuous tuning range of most CW lasers is only about 30 GHz; there is then a jump to a new tuning band.

The saturation phenomenon described in the previous section can be used to stabilize a reference laser. A gas cell with an absorption line at a frequency ν_0 somewhere within the laser gain curve is placed in the laser cavity. If the laser is operating at a frequency $\nu_0 - \Delta\nu$, the light is absorbed by atoms travelling towards the incident radiation, which see it as blue shifted into resonance, and a hole is burned in the population distribution at the velocity v_1 given by equation 8.5; but the cavity contains a wave of equal amplitude travelling in the opposite direction, and this is absorbed by atoms of equal and opposite speed, so another hole is burned at the velocity $-v_x$, as shown in Fig. 8.6(a). When the laser is tuned (by adjusting the cavity length) to exactly ν_0 the two holes coalesce (Fig. 8.6(b)), and the gain of the cavity increases as shown in Fig. 8.7. The pip on the Doppler distribution has the homogeneous width of the absorbing transition and is often called the 'inverted Lamb dip'. The true Lamb dip refers to an analogous saturation effect in the lasing medium itself: the laser gain has a sharp dip at ν_0 because fewer atoms can contribute to the stimulated emission when both left- and right-travelling waves are using the same group of atoms. Any atom or molecule with sharp absorption lines in the appropriate spectral region can be used for locking on the inverted Lamb dip in this way; methane (CH_4) is the most useful in the near infra-red and iodine (I_2) in the visible region. The

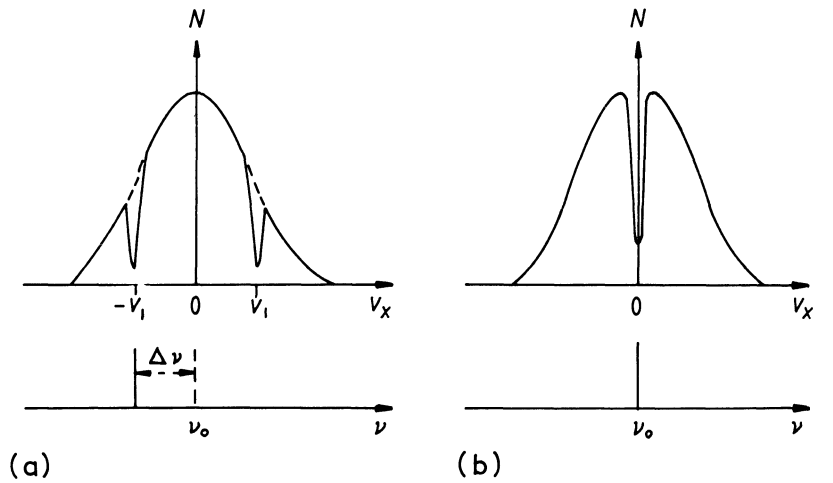

Figure 8.6 Laser stabilization by saturation. For detuning by $\Delta\nu$, 'holes' are burned at $\pm v_1$, as shown in (a). When $\Delta\nu = 0$, as in (b), the 'holes' coalesce.

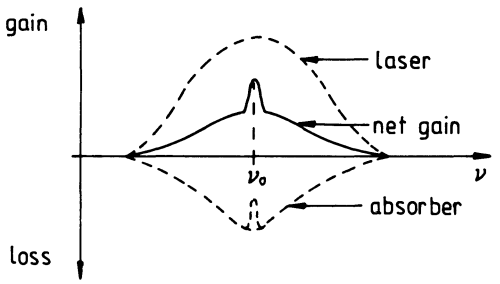

Figure 8.7 Frequency locking. The laser gain curve is reduced by the loss from the absorber, but the hole in the absorption (see Fig. 8.6 (b)) gives a reference pip on the net gain curve.

long-term stability and reproducibility of such lasers can be as good as 5 and 50 kHz, respectively – well below the actual molecular linewidth.

The inherent stability of about $1:10^{11}$ of Lamb-dip stabilized lasers makes them obvious candidates for frequency and/or wavelength standards in the infra-red and visible regions. The fundamental frequency standard from the cesium hyperfine structure at 9 GHz (section 9.12) is accurate to $1:10^{13}$. This cannot be compared directly with a visible or near infra-red frequency at a few hundred THz, but frequency mixing along a multistage chain using first klystrons and then several different infra-red stabilized lasers has transferred the frequency standard to stabilized He–Ne lasers at 3.39 μm with an accuracy of a few parts in 10^{11}, and on to the I_2-stabilized He–Ne line at 633 nm at about $1:10^{10}$. With the velocity of light now defined as a constant (section 1.7) these frequency standards also serve as wavelength standards to the same accuracy, an improvement of an order of magnitude over the wavelength standard provided by the incoherent krypton lamp ($4:10^9$). The accuracy with which wavelengths can be compared interferometrically, a few parts in 10^{10}, thus matches the visible frequency standards [11, 12].

Temporal resolution, obtained from very short laser pulses, offers the possibility of measuring lifetimes of excited states and following collision rates and fast chemical and biological reactions. Pulses of a few ns can be obtained by inserting in the laser cavity a Pockel cell, which is essentially a variable polarizer. This is switched 'off' until pumping of the laser transition is complete; when it is switched 'on' the gain increases so rapidly that the level population reaches equilibrium in a few ns. This so-called Q-switching is used to produce very high powers as well as short pulses.

Much shorter pulses can be generated by mode-locking. A cavity of length L without additional tuning can support a number of modes, as shown in Fig. 8.2, the wavelengths of which are defined by integral values of m in the relation $m\lambda = 2L$. If the refractive index of the cavity is chosen so that the *frequency* separations of adjacent modes are all equal, there is a constant beat frequency at

the mode separation, $c/2L$ Hz – i.e. all the modes come into phase at time intervals $\Delta t = 2L/c$. They stay in phase only for a time δt inversely proportional to the frequency band over which they are spread – the full width of the gain curve in Fig. 8.2. If this is 10^{12} Hz, which is 1 nm in the visible region, δt is 1 ps. The spectral bandwidth of these picosecond pulses is still that of the relatively narrow cavity mode bandwidth. The bandwidth theorem is not violated, despite appearances to the contrary, because it is the coherent superposition of many such pulses that shortens the pulse length. Pulses of about 0.01 ps have actually been obtained; this is almost unbelievably short when one remembers that it corresponds to only a few cycles of the optical frequency.

8.5 Laser absorption and excitation experiments [1–7, 14–21, 28, 30]

Atomic and molecular absorption lines can be recorded without the need for a spectrometer by scanning a tunable laser across the line profile. There are several important advantages over conventional spectroscopy with a broad-band background source and a spectrometer. First, higher spectral resolution is usually attainable with a laser. Secondly, the low divergence makes it relatively easy to use long absorption paths, including multipass cells, to observe weak transitions. Thirdly, the high photon flux populates the upper level of the transition so efficiently that absorption can be detected by the subsequent fluorescence even when the transition is weak or the number of absorbing atoms is small. Indeed, laser-induced fluorescence is usually the most sensitive of all analytical techniques for detecting trace quantities of elements. It will be seen in section 8.7 that fluorescence from single atoms or ions confined in traps can be detected.

The high flux density attainable from tunable lasers can be used to perform efficient stepwise excitation, as shown schematically in Fig. 8.8(a). The intermediate level, level k in the figure, may be populated either by a discharge or by a tunable laser; in the latter case the second step generally requires a second tunable laser, and the final level has the same parity as the ground state.

With sufficient laser intensity two photon transitions become possible, as shown in Fig. 8.8(b). The intermediate level is in this case a virtual level of opposite parity to both the initial and final states. The transition probability is found from second-order perturbation theory to be proportional to the square of the laser intensity and to the product of the transition moments $i \rightarrow n$ and $n \rightarrow f$ (cf. equation 2.32) summed over all real states n:

$$P_{if} \propto I^2 \sum_n \frac{|\langle i|\boldsymbol{er}|n\rangle\langle n|\boldsymbol{er}|f\rangle|^2}{(\omega - \omega_{ni})} \tag{8.6}$$

We shall meet two-photon processes again in section 9.10 in connection with Raman spectroscopy. It is evident from equation 8.6 that the probability of the two-photon transition is much enhanced if the intermediate level is close to a

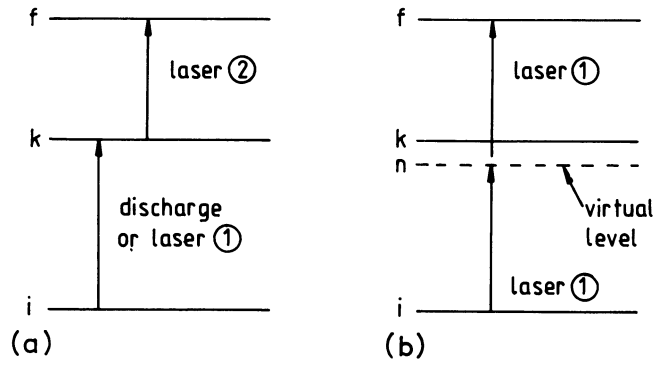

Figure 8.8 (a) Stepwise excitation and (b) two-photon excitation. Level k in both cases is a real level. The two-photon probability is enhanced when the virtual level n (shown broken) is close to a real level k.

real level. P_{if} does not, of course, actually become infinite when $\omega = \omega_{ni}$ because there is a damping term that has not been included in equation 8.6. Two-photon transitions are important in several respects. They connect the ground state with states of the same parity – for example, s–s and s–d transitions are allowed; transitions in the ultra-violet can be probed with visible region tunable lasers; and two-photon Doppler-free spectroscopy (see next section) has proved a particularly powerful technique.

There is no reason for stopping at two photons; multiphoton excitation and ionization have proved entirely feasible. The three-photon transition probability depends on the cube of the laser intensity and on terms analogous to equation 8.6, but with the product of three matrix elements in the numerator and two frequency differences in the denominator. Again the cross-section is much enhanced whenever one of the virtual levels coincides with a real level.

Efficient selective population of excited levels, whether by one or by more photons, has made possible a wide variety of experiments, of which the following must be regarded as examples, not a complete list. The absorption spectrum from the populated excited level to higher levels or to the continuum can be probed, either by a second laser or by conventional methods. The lifetime of the level can be measured, as further discussed in Chapter 12. If the state is fairly long-lived its structure can be probed by magnetic resonance beam deflection methods (Chapter 9). The various double resonance and level-crossing experiments described in sections 9.8 and 9.9 rely on optical excitation, and their scope and sensitivity have been greatly increased by the use of tunable lasers instead of conventional light sources.

Selective excitation is also important as a method of isotope shift measurement and isotope separation. If the laser bandwidth is smaller than the isotope shift (section 2.11) in the transition concerned, one isotope at a time can be excited, or at least preferentially excited. If a second photon from the same laser,

or from a second broadband laser, photo-ionizes the atom, as shown in Fig. 8.9(a), the ion can be extracted with an electric field by virtue of its positive charge. Actually, it is not necessary for the second laser to take the atom all the way to its ionization limit: if one of the closely packed high-lying states is populated, as shown in Fig. 8.9(b), the ionization process can be completed by applying an electric field, a process known as field ionization. The strength of field required depends on the energy gap to be bridged. The selective excitation/field ionization method is extremely sensitive, as in principle every atom capable of absorbing a photon of the particular frequency to which the laser is tuned can be ionized and detected. It is actually possible to detect one atom against a background of about 10^{19} atoms or isotopes of slightly different resonance frequency. At the other end of the sensitivity scale the method has acquired enormous industrial (and strategic) importance in the large-scale separation of uranium isotopes.

A (literally) more academic method of isotope separation is that of beam deflection. If an atomic beam collimated by a series of slits is crossed by a laser beam tuned to a resonance transition, the atoms acquire a transverse momentum by virtue of the photons absorbed. This momentum is only $h\nu/c$ per absorption – of order $10^{-27}\,\mathrm{J\,s\,m^{-1}}$ in the visible region – compared with a momentum in the beam direction four or five orders of magnitude higher. However, if the lifetime of the excited state is short (10 ns being typical for a resonance line), an atom can absorb and re-emit a hundred or more photons during its passage across the beam, and insofar as the re-emission is spontaneous it does not, on average, change the momentum. Deflections sufficient to include or exclude particular isotopes at the exit slit can be attained in this way.

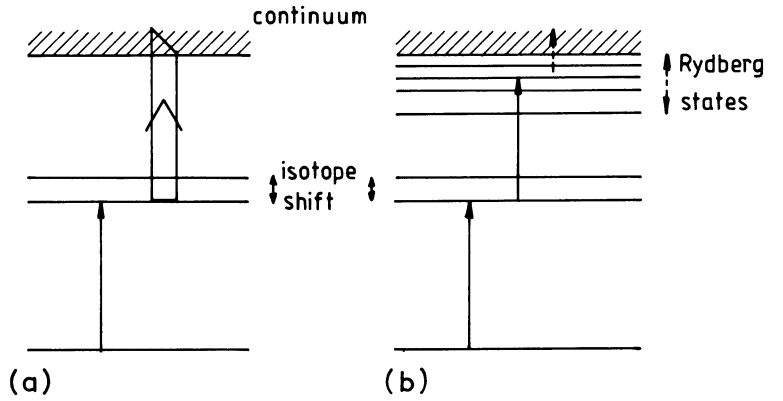

Figure 8.9 Selective ionization. A narrow-band laser is tuned to the resonance frequency of a particular isotope. The excited atoms can then be ionized by a broadband laser, as in (a), or further excited to a high n-state, as in (b), from which field ionization is possible (see text).

The properties of the states of high n shown in Fig. 8.9(b) – so-called Rydberg states – are themselves of great interest. Since the mean distance of an electron from the nucleus increases with n^2, a highly excited electron sees the nucleus and core electrons almost as a point charge $+e$, and the atom becomes increasingly hydrogen-like. A number of strange effects result both from its macroscopic 'size' (for $n = 100$, the 'diameter' of the atom is 1 μm) and from the weakness of the electron–nuclear binding energy, which decreases as n^2 and is therefore only about 1 meV for $n = 100$. Relatively weak external fields appear 'strong' to Rydberg atoms, allowing the study of quadratic Zeeman and higher-order field effects, the mixing of states of different n value, and so on. The large atomic radius gives rise to a large electric dipole moment, with interesting consequences for interactions with other atoms and for collisional cross-sections.

Another feature of Rydberg atoms of high l-value, excited by multiphoton absorption, is that they cannot decay directly to the ground state because of the Δl selection rule, and as transitions to nearby high n-states are in the microwave frequency range, where spontaneous transition probabilities are very small, they actually have rather long lifetimes. By introducing such atoms into a microwave cavity it is possible to conduct very fundamental studies of the interactions between a few atoms and a few photons.

8.6 Doppler-free spectroscopy [2, 5–7, 13–16, 30]

In Doppler-free spectroscopy the smearing out of the homogeneous line profile by the velocity spread of the absorbing atoms is avoided by selecting a subset of atoms with a particular line-of-sight velocity – normally zero velocity. There are various ways of achieving this, but the principle can be explained in the context of the original 'hole burning' method.

As shown in Fig. 8.5, a sufficiently powerful narrow-band laser tuned to a frequency $\Delta\nu$ from the centre ν_0 of an absorption line can burn a hole through the absorption profile by saturating the transition – i.e. equilibrating the upper and lower state populations – for all atoms with the particular value of v_x given by equation 8.5. A probe laser tuned to the same frequency finds the gas transparent if it is going in the same $(+x)$ direction, but if its direction is reversed the atoms it probes are those with velocity $-v_x$, and the gas no longer appears transparent. Only if both lasers are tuned to ν_0 does the probe see the same set of atoms as the bleaching laser – those with zero v_x. In practice a single laser with a beamsplitter and appropriate mirror arrangement forms both bleaching and probe beams, as shown in Fig. 8.10. The laser is scanned through the line profile, and the probe detector registers a sharp peak at line centre. This phenomenon is essentially the same as the inverted Lamb dip locking described in section 8.4.

The technique is made much more sensitive by modulating the bleaching laser so that the line centre is detected by the appearance of modulation in the probe detector. This reduces the effects of background and scattered light to the extent

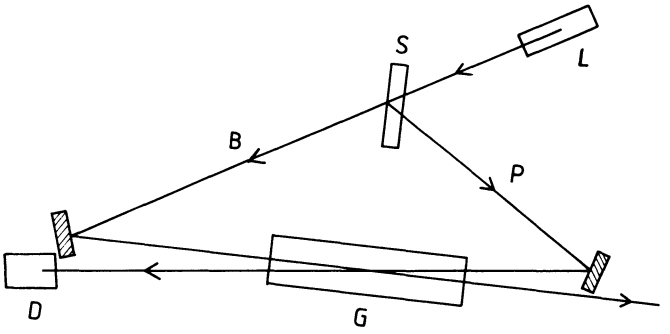

Figure 8.10 Doppler-free spectroscopy by saturation. The laser beam L is divided at the beamsplitter S into an intense bleaching beam B and a weak probe beam P, which pass in opposite directions through the gas G. Tuning to line centre is registered by an increase in the probe beam signal at the detector D.

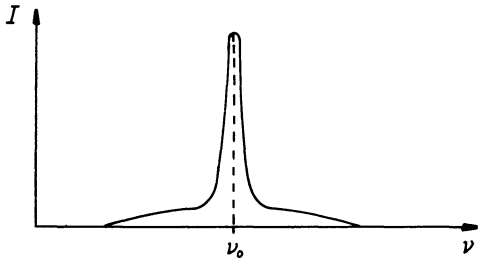

Figure 8.11 Doppler-free signal for an arrangement such as that of Fig. 8.10. The Doppler-free spike rises from a base of the full Doppler width.

that it is possible to detect, for example, Doppler-free transitions in a discharge. The hole in the Doppler distribution is constantly refilled by velocity changing collisions in the gas, so the Doppler-free peak actually rises from a platform of the full Doppler width (Fig. 8.11).

Various modifications of the basic absorption saturation technique have been devised to improve the sensitivity and reduce noise so as to detect weaker signals. If the bleaching laser is circularly polarized, the atoms are oriented in the manner illustrated in Fig. 8.4, and the Doppler-free peak can be detected by a change in polarization of the (plane-polarized) probe laser, a small change in polarization being more easily detected than a small change in intensity. Modulated fluorescence can also be more sensitively detected than modulated absorption. Further clever tricks can be played by modulating both beams at different frequencies or by modulating the polarization.

The ultimate bandwidth limit is the natural linewidth, which is of order 10 MHz for a strong line and a few kHz for 'forbidden' lines (section 10.2). This

has been approached, or even reached, in some experiments. The difficulties of working to such limits are increased by the necessity of using low gas pressures and reasonably low laser intensities to avoid pressure and power broadening, respectively (sections 10.4 and 8.7). Probably the most fundamental measurement made by saturation absorption spectroscopy has been the absolute wavelengths of the fine structure components of the Balmer lines of hydrogen and deuterium (section 2.7): these have yielded a value of the Rydberg constant accurate to about three parts in 10^9, an improvement of a couple of orders of magnitude on Doppler-limited measurements.

The technique of two-photon Doppler-free spectroscopy is essentially a complementary one, applicable to transitions forbidden to single-photon spectroscopy as discussed in the last section. In this case the laser beam is split into two oppositely propagating beams of equal intensity, and the Doppler-free transition occurs for atoms capable of absorbing one photon from each beam, which means zero velocity in the beam direction. Again there is a Doppler-broadened plateau corresponding to absorption of two photons from one beam.

The most fundamental application of this technique is again to the fine structure of the hydrogen atom – in this case to the two photon $1s^2S - 2s^2S$ transition. At densities sufficiently low that collisional transfers from $2s^2S$ to the almost degenerate $2p^2P$ level are rare, the former is metastable, and the intrinsic natural linewidth is only a few Hz. By the Balmer formula, the frequency of the 1s–2s transition is $(1-\frac{1}{4})$ Rydbergs, $=\frac{3}{4}R_H$, corresponding to two photons at 243 nm. Now the frequency of the Balmer β-line, $n=2$–$n=4$, is given by $\nu=(\frac{1}{4}-\frac{1}{16})R_H=\frac{3}{16}R_H$, corresponding to 486 nm. A laser tuned near H_β can therefore serve as the fundamental for frequency doubling to produce the 243 nm radiation for the two-photon transition, and the difference from an exact multiple of four between the two resonances measures the Lamb shift of the 1s state of hydrogen, $\simeq 8150$ MHz, a most important quantity for comparison with quantum electrodynamical theory (section 2.7). The final limiting accuracy on this measurement seems likely to be a few tens of kHz.

The high powers required for two-photon spectroscopy are much more easily achieved with pulsed than with CW lasers. The limiting resolution is then often the frequency spread $1/\tau$ associated with the finite pulse length τ. In the case of CW lasers a similar limit is imposed by the time of flight of an atom through the laser beam. The possibility of using optical interference effects to get around this limitation has been demonstrated for both spatial and temporal separation of the two absorptions. In the first case, applied to an atomic beam, an atom experiences two coherent laser fields separated by the time of flight T between them; in the second case T is the time delay between two coherent laser pulses (an incident pulse and its reflection from a mirror). In both cases the resonance manifests itself as a two-beam interference pattern in the frequency domain: a set of interference fringes of separation $1/T$ limited by an envelope of width $1/\tau$. The effect can be thought of as a temporal equivalent to a Young's slits set-up, with τ replacing the slit widths and T their separation. An analogous device was used

by Ramsey to improve resolution in molecular beam magnetic resonance experiments (section 9.6), and for this reason the name 'optical Ramsey fringes' is often used.

8.7 High intensity effects, cooling and trapping [2, 22–28, 30]

8.7.1 HIGH INTENSITY LASER FIELDS

The interaction of high intensity coherent electromagnetic fields with atoms and molecules has been studied for many years in radiofrequency spectroscopy. In the optical region such interactions were masked by Doppler broadening until the advent of high-power tunable lasers, on one hand, and Doppler-free techniques on the other.

The two important effects at and near resonance with an absorption line are both well known in radiofrequency spectroscopy: power broadening, and the Autler–Townes, or A. C. Stark, effect, which splits each of the energy levels involved in the transition. Both effects are due to the interaction energy V_{12} between the electric field of the laser and the electric dipole transition moment p_{12}. Assuming for simplicity plane-polarized radiation with an electric field $E_0 \cos \omega t$ in the x-direction, then $p_{12} = \langle 2|ex|1\rangle$ in the notation of section 2.8, and E_0 is related to the laser flux density I by equations 4.7 and 4.8,

$$I = \tfrac{1}{2} \varepsilon_0 c |E_0|^2 \Rightarrow E_0 = 27\sqrt{I} \tag{8.7}$$

with E_0 in V m^{-1} and I in W m^{-2}. The interaction is characterized by the Rabi frequency Ω defined by

$$\Omega = \frac{V_{12}}{\hbar} = \frac{E_0 |p_{12}|}{\hbar} \tag{8.8}$$

The interaction V_{12} is identical with that treated by time-dependent perturbation theory in section 2.8, but we are now envisaging it to be too large to treat as a perturbation. If the transition is saturated in the way described earlier in this chapter it is clearly not valid to treat the lower-state population as unaffected by the radiation.

The power broadening is seen when the laser is detuned from exact resonance. Suppose, ignoring Doppler broadening for the moment, we measure the total fluorescence (integrated over the linewidth) as a function of detuning Δ. The fluorescence signal is proportional to the total power absorbed. As Δ is increased the scattered power decreases, following the Lorentzian distribution of section 10.2, but the width of this detuning curve is proportional to $\sqrt{(\gamma^2 + \Omega^2)}$ instead of to the normal homogeneous linewidth γ. For large Ω the width increases linearly with Ω. If Doppler broadening is significant, the detuning Lorentzian has to be convoluted with the appropriate Gaussian, and the power broadening is only apparent if Ω is comparable with the Doppler width.

The second effect appears if a narrow-band laser is tuned to resonance: each level of the transition is split into two levels separated by $\hbar\Omega$. This splitting can be thought of as a linear Stark effect, proportional to the amplitude of the laser field. Alternatively, the atom and field together can be treated as a 'dressed atom' with two states, $E_1+(n)$ photons and $E_2+(n-1)$ photons. At resonance, when the photon energy $\hbar\omega_L$ is equal to E_2-E_1, the two states are degenerate in zero order, but the interaction between them acts as a perturbation in the manner first discussed in connection with the helium atom in section 2.6, splitting them by the amount V_{12}. Thus, a transition from either level to a third level E_3 has two components separated by $\hbar\Omega$ as shown in Fig. 8.12, while the fluorescence shows three components at ω_0 and $\omega_0 \pm \Omega$. The condition for observation of the splitting is that dephasing effects must be small, which means in practice that both the homogeneous and inhomogeneous linewidths and the laser bandwidth must be small compared with Ω. To give an order of magnitude, Ω for a flux density of 1 MW cm^{-2} combined with a typical resonance line strength is of order 10^{11} rad s^{-1}, or 10–20 GHz, compared with $\sim$1 GHz for a typical Doppler width.

A calculation of the probability of finding the atom in one of the levels E_1 or E_2 shows that it is cycled between them with the Rabi frequency Ω. This suggests a semi-classical interpretation of the three-component fluorescence spectrum: the fluorescence, following the population of level 2, is modulated at frequency Ω and thus gives rise to a fundamental at ω_0 with two sidebands at $\pm\Omega$.

When the laser is detuned slightly from resonance the three-component pattern becomes asymmetric. However, we cannot pursue these high intensity

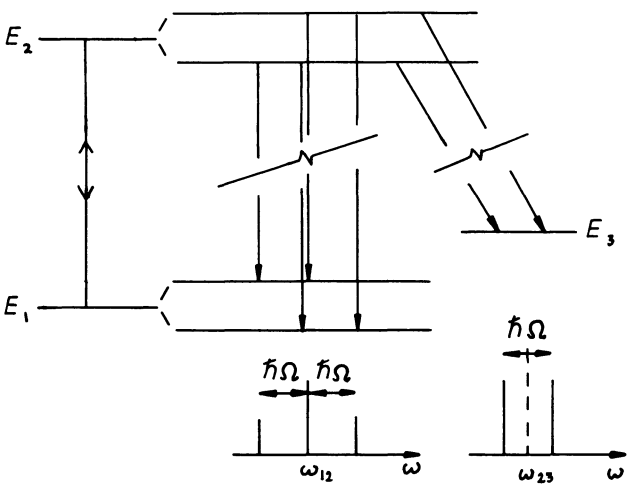

Figure 8.12 A. C. Stark (Autler–Townes) effect. Interaction between laser field and atom splits the pumped levels E_1 and E_2 into doublets of separation $\hbar\Omega$. Resonance fluorescence appears as a triplet. A non-resonant transition to a third level E_3 is a doublet.

effects any further here; relevant references are given at the beginning of the subsection.

8.7.2 LASER COOLING

The photon momentum transfer process described in section 8.5 in connection with isotope separation can be used to slow down and thus cool atoms or ions in an atomic beam or a 'trap'. Radiation from a laser tuned to the long-wavelength side of a resonance line is absorbed by atoms travelling towards the laser, since they see the radiation as blue-shifted. Their momentum in this direction is decreased with each absorption by $h\nu_L/c$, where ν_L is the laser frequency. Subsequent re-emission is in a random direction and does not affect the average momentum. The average kinetic energy loss per absorption is $h(\nu_0 - \nu_L)$, and about 10^4 absorptions are needed to get from thermal velocities to a few K. For a strong transition the lifetime of the excited state is of order 10 ns, and the required recycling can be achieved in a few ms. There are some obvious practical difficulties to be overcome. As the atoms cool, the laser needs to be tuned closer to resonance, which can be done either by sweeping the laser frequency ν_L or by Zeeman-shifting the resonance frequency ν_0 with an external magnetic field. Secondly, the laser reduces momentum in one direction only, and the random recoil momentum from spontaneous emission is tantamount to heating. In principle six lasers are required to cool free atoms efficiently. The cooling of ions whose motion is constrained by electric and magnetic fields – i.e. trapped ions – requires only one laser.

Experiments on cooled atoms or ions differ from the Doppler-free experiments described in the previous section in that all atoms or ions participate, instead of just a subset with a particular velocity. Moreover, the second-order transverse Doppler shift is also eliminated. The desirability of trapping neutral atoms provides an additional incentive for cooling, since all such traps are too shallow to contain fast atoms.

8.7.3 ION TRAPS

Ion traps use some combination of electric, or electric and magnetic, fields (the configurations of which will not be discussed here) to create a potential minimum of several volts over a small region of space, such that a finite number of ions, from one to several thousand, can be trapped there for minutes or even hours. Laser cooling of the trapped ions, using a single laser, to below 1 K has been accomplished. It is possible to make measurements of high precision on isolated and almost stationary ions, free of first- and second-order Doppler effects and of pressure and transit time broadening. A resolution of up to $1:10^{14}$ may ultimately be attainable, and it seems likely that trapped ions will provide the frequency standards of the future.

The measurements themselves may involve any of the laser excitation and

radiofrequency techniques described in this and the next chapter. The principal applications at the time of writing have been to hyperfine structure (yielding an improved value for the ratio of electron to proton mass), and to Lamb shifts and isotope shifts. The sensitivity can be made so high that fluorescence from single ions can be detected. This makes possible measurements on minute quantities of rare isotopes, the shifts of which yield important data on nuclear shapes and distortion factors.

8.7.4 TRAPPING OF NEUTRAL ATOMS

A method of trapping neutral particles is obviously much to be desired, and at the time of writing some progress has been made towards the implementation of 'light traps'. These exploit the fact that a dipole in an inhomogeneous field experiences a force in the direction of the field gradient. The principle is the same as that used in the Stern Gerlach experiment and in the magnetic deflection arrangements described in Chapter 9, but in this case the field is the electric vector of the laser radiation and the dipole is induced by the electric field. The deflecting force $\boldsymbol{F}$ is given by

$$\boldsymbol{F} = -\tfrac{1}{2}\alpha \nabla |E|^2$$

where α is the polarization and $\boldsymbol{E}$ is the electric field vector, related to the flux density in the laser by equation 8.7. The sign of the polarization depends on whether the laser is tuned above or below resonance (for the same reason that the phase of a driven oscillator changes sign at resonance), and its magnitude depends on the amount of detuning Δ and on the intensity of the laser. For $I \gg I_s$, where I_s is the saturation flux density of equation 8.4, it can be shown that the force is given by

$$\boldsymbol{F} = h\frac{\Delta}{I}\nabla I$$

the direction being such that for $\nu_L < \nu_0$ ('red detuning') the atom is pushed in the direction of increasing I. An intense laser beam focused to a narrow 'waist' therefore constitutes a one-dimensional trap for neutral atoms. In principle three such beams would complete a trap, but there are in fact problems of stability associated with such an arrangement, and the depth of the trap is in any case very small.

References

GENERAL FURTHER READING

1. Corney, A. (1977) *Atomic and Laser Spectroscopy*, Oxford University Press.
2. Loudon, R. (1973) *The Quantum Theory of Light*, Oxford University Press.
3. Mooradian, A., Jaeger, T. and Stokseth, P. (eds) (1976) *Tunable Lasers and Applications* (*Springer Series in Optical Sciences*, Vol. 3), Springer-Verlag, Berlin.

4. Schafer, E.P. (ed.) (1973) *Dye Lasers* (*Topics in Applied Physics*, Vol. 1), Springer-Verlag, Berlin.
5. Series, G.W. (1984) Laser spectroscopy, *Contemp. Phys.*, **25**, 1.
6. Walther, H. (1976) *Laser Spectroscopy of Atoms and Molecules*, Springer-Verlag, Berlin.

In this general category we also put the two 'Nobel prize articles'

7. Bloembergen, N. (1982) Non-linear optics and spectroscopy, *Rev. Mod. Phys.*, **54**, 685; and Schawlow, A.L. (1982) Spectroscopy in a new light, *ibid.*, 697.

DIODE LASERS AND ULTRA-VIOLET LASERS

8. Camparo, J.C. (1985) The diode laser in atomic physics, *Contemp. Phys.*, **26**, 241.
9. Jamroz, W. and Stoicheff, B.P. (1983) Generation of tunable coherent VUV radiation *Progress in Optics*, Vol. 20, North-Holland, Amsterdam.
10. Sobelmann, I.I. and Vinogradov, A.V. (1985) On the problem of extreme UV and X-ray lasers, *Adv. Atomic Molec. Phys.*, **20**, 327.

FREQUENCY STANDARDS

11. Baird, K.M. (1983) Frequency measurement of optical radiation, *Phys. Today*, **37**, 52.
12. Petley, B.W. (1983) *The Fundamental Physical Constants and the Frontiers of Measurement*, Hilger, Bristol.

HIGH-RESOLUTION AND DOPPLER-FREE SPECTROSCOPY

13. Knight, P.L. (1981) Laser spectroscopy beyond the natural linewidth, *Comments Atomic Molec. Phys.*, **10**, 241.
14. Lehmann, J.C. (1978) Recent developments in the spectra of small molecules, *Rep. Prog. Phys.*, **41**, 609.
15. Shimoda, K. (ed.) (1976) *High Resolution Laser Spectroscopy* (*Topics in Applied Physics*, Vol. 13), Springer-Verlag, Berlin.
16. Thompson, R.C. (1985) High resolution laser spectroscopy of atomic systems, *Rep. Prog. Phys.*, **48**, 531.

MULTIPHOTON TRANSITIONS AND RYDBERG ATOMS

17. Stenholm, S. (1979) Laser-induced multiphoton transitions, *Contemp. Phys.*, **20**, 37.

Four review-type articles have been published in *Advances in Atomic and Molecular Physics* (D.R. Bates and B. Bederson eds):

18. Edelstein, S.A. and Gallagher, T.F. (1978) Rydberg atoms, Vol. 14, p. 365.
19. Feneuille, S. and Jacquinot, P. (1981) Atomic Rydberg states, Vol. 17, p. 99.
20. Gallas, J.A.C., Leuchs, G., Walther, H. and Figger, H. (1985) Rydberg atoms: high resolution spectroscopy and radiation interaction, Vol. 20, p. 414.
21. Haroche, S. and Raimond, J.M. (1985) Radiative properties of Rydberg states and resonant cavities, Vol. 20, p. 350.

HIGH LASER FIELDS

22. Knight, P.L. and Milonni, P.W. (1980) The Rabi frequency in optical spectra, *Phys. Rep.*, **66**, 21.

23. Mollow, B.R. (1981) Theory of intensity-dependent resonant light scattering and fluorescence, *Progress in Optics*, Vol. 19.

LASER COOLING AND TRAPPING

24. Letokhov, V.S. and Minogin, V.G. (1981) Laser radiation pressure on free atoms, *Phys. Rep.*, **73**, 1.
25. Stenholm, S. (1986) The semiclassical theory of laser cooling, *Rev. Mod. Phys.*, **58**, 699.
26. Werth, G. (1985) Trapped ions, *Contemp. Phys.*, **26**, 241.
27. Wineland, D.J. and Itano, W.M. (1983) High resolution spectroscopy of stored ions, *Adv. Atomic Molec. Phys.*, **19**, 135.

MISCELLANEOUS

28. Letokhov, V.S. (1981) Is it possible to detect very rare atoms by means of laser radiation? *Comments Atomic Molec. Phys.*, **10**, 257.
29. New, G.H.C. (1983) The generation of ultrashort laser pulses, *Rep. Prog. Phys.*, **46**, 877.
30. "*State-of-the art*" *articles* are to be found in the Springer Series in Optical Sciences *Laser Spectroscopy*, published in alternate years by Springer-Verlag, notably Vols III, IV, V and VI for 1977, 1979, 1981 and 1983, respectively.

This list should make an adequate starting point for further reading, but it is undoubtedly incomplete, even at the time of writing. The review journals cited in this list should be consulted for relevant articles published after 1985–86.

NINE

Other spectroscopic methods: microwave, radiofrequency, level-crossing and Raman

9.1 Introduction

Microwave and radiofrequency spectroscopy measure directly relatively small energy separations, between about $10\,cm^{-1}$ and a few thousandths cm^{-1}, or 300 GHz to 30 MHz. The short wavelength end of this range (1 mm $< \lambda <$ 10 cm, say) comprises pure rotational transitions of molecules, Zeeman and Stark splittings, and the hyperfine structure of heavy atoms. Towards longer wavelengths (lower frequencies) one can measure the hyperfine structure of light atoms, the fine structure of hydrogen-like ions and Zeeman splitting from the nuclear magnetic moment.

All these effects can be observed as small perturbations (shifts or splittings) of transitions in the optical region – by resolving, for example, the rotational structure of molecular bands or the hyperfine structure of atomic lines in the visible region – but they then appear as very small differences in very large quantities. A typical hyperfine splitting in a light atom, a few hundredths cm^{-1}, represents a difference of one part in 5×10^5 in the visible transitions. This is very close to the limit set by the Doppler width in conventional measurements – a limit that can in suitable cases be overcome by the Doppler-free methods of section 8.6. Since the Doppler width is a constant fraction of the frequency (section 10.3), the alternative is to measure the splittings directly at very much lower frequencies. The methods described in this chapter exploit this alternative.

As shown in Chapters 2 and 3, radiative transitions between two energy levels depend on the existence of a transition moment, which may be electric dipole, magnetic dipole, electric quadrupole or higher order. This requirement limits the techniques that are applicable to any particular problem. For example, most heteronuclear molecules have a permanent electric dipole moment; transitions between hyperfine or rotational levels of the same electron configuration are possible, and the molecule has a pure rotational absorption spectrum. Atoms

and homonuclear molecules have no permanent electric dipole moment and can undergo electric dipole transitions only between different electronic states of opposite parity; but an atom with non-zero resultant angular momentum $\boldsymbol{J}$ can make radiative transitions between hyperfine levels of a single configuration by virtue of its permanent magnetic dipole moment.

Also included in this chapter are two techniques that can be used to determine small energy level separations by the scattering of visible light rather than the absorption of microwave or radiofrequency radiation; these are level-crossing and Raman spectroscopy. The first uses polarization changes to determine hyperfine structure, and the second uses inelastic scattering to determine energy differences for which direct transitions are forbidden – vibrational and rotational structure of homonuclear molecules, for example.

In this chapter we consider first straightforward electric dipole absorption of microwaves by free molecules (i.e. gases). Next come the wide class of magnetic resonance techniques in which the frequency is tuned to the energy differences between magnetic sub-levels of an atom or molecule in an external magnetic field. Magnetic resonance techniques fall into two categories. In the first the existence of a resonance is detected in a straightforward fashion by measurement of the absorbed energy, either as electron spin resonance (ESR) or as nuclear magnetic resonance (NMR); these techniques are effectively limited to liquids and solids and are discussed in outline only. In the second category come experiments in which magnetic resonance is detected in free atoms and molecules by beam deflection or double resonance; these are described in rather more detail, together with the closely related level-crossing technique. Two sections outlining Raman spectroscopy, both spontaneous and coherently driven (CARS), conclude the chapter.

9.2 Microwave spectroscopy [2, 4]

The microwave region may be taken, in round figures, as 10^6–10^3 MHz, or 30–0.03 cm^{-1}, the high-frequency limit being set by the range of klystron oscillator harmonics. Microwave spectroscopy of gases is confined to absorption because the emission spectra from laboratory sources is too weak to be of practical importance. There is no need for dispersing elements in microwave spectroscopy: the source and detector are tuned together across the relevant frequency range while the signal is recorded as a function of frequency. The effects studied are primarily rotational transitions of molecules, in which electric quadrupole effects may appear as a fine structure, and Stark shifts in an external electric field. As seen in Chapter 3, the rotational constants of most molecules are of order 1 cm^{-1}, near the middle of the range given above, but molecules such as the diatomic hydrides have such large rotational constants that their rotational spectra extend over the infra-red boundary. Homonuclear and symmetric-top molecules, which have no permanent electric dipole moment, have no

pure rotational spectrum. The classic experiment of Lamb and Retherford on the fine structure of hydrogen also falls into the category of electric dipole transitions in the microwave region.

Figure 9.1 shows schematically a typical layout for microwave absorption. The source is normally a klystron oscillator, and the radiation from it passes through a waveguide of suitable dimensions, which acts as the absorption cell, to a crystal detector whose output is proportional to the incident power. Various methods of modulation, phase sensitive detection, heterodyning, etc., may be used to improve the signal-to-noise ratio that would be obtained from a simple d.c. scan.

The accuracy with which microwave frequencies can be measured is of order $1:10^7$ but, as in optical spectroscopy, the precision with which the frequency of a transition can be determined is usually limited by the intrinsic linewidth. Of the line-broadening mechanisms discussed in Chapter 10, pressure broadening is the most important in the microwave region. At 10^{-2} torr the full width at half-maximum (FWHM) of a pressure-broadened line is about 0.1 MHz. With a frequency of order 30 000 MHz (1 cm wavelength) one can expect a resolving power in the range 10^5–10^6. It is not possible to reduce this appreciably by working at low gas pressures because of the onset of power saturation broadening. This occurs when molecules are raised to the upper state by absorption of radiation more quickly than they can be returned to the lower state by collisions. As the power is raised, the absorption at the peak frequency saturates first, while the wing absorption continues to grow, resulting in a broader line. Spontaneous radiative decay plays an insignificant part in the relaxation process

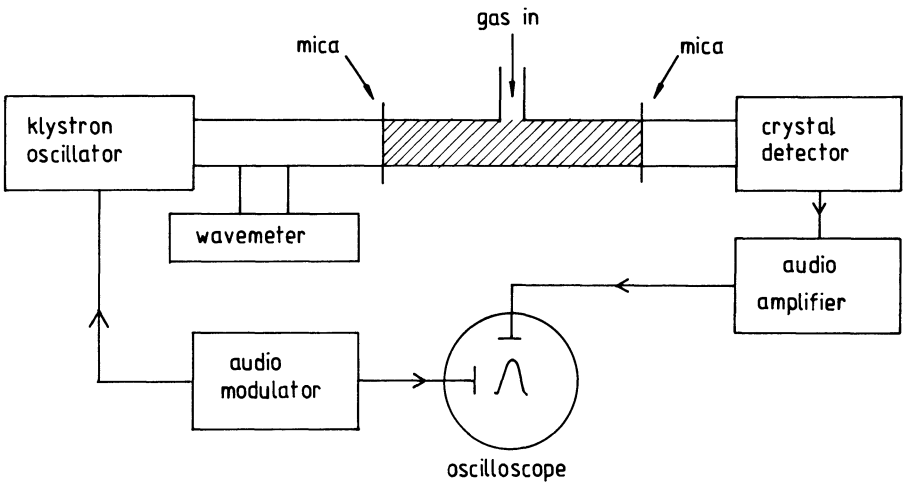

Figure 9.1 Schematic arrangement for microwave absorption spectroscopy. A section of waveguide isolated by mica windows acts as an absorption cell. The source is swept at audio-frequency over a small frequency range.

at these frequencies (see section 4.5 for the ν^3 factor). With power input of the order of a mW, the power saturation and pressure broadening at 10^{-2} torr are of the same order of magnitude. Since one goes up as the other goes down when the pressure is changed, there remains an irreducible linewidth of the order of 0.1 MHz.

9.3 Microwave spectroscopy of gases: the principal applications

9.3.1 ROTATIONAL CONSTANTS

The rotational energy of a diatomic or linear polyatomic molecule is given by equation 3.16:

$$F(J)=B_v\,J(J+1)-D_v\,J^2(J+1)^2$$

where B is inversely proportional to the moment of inertia, D is a small correction for centrifugal stretching, and both B and D are functions of the vibrational quantum number v as well as the electronic state. The selection rule for electric dipole radiation is $\Delta J=\pm 1$, so absorption can occur for wavenumbers σ given by

$$\sigma=2B(J+1)-4D(J+1)^3 \quad \text{cm}^{-1}$$

where B and D are in cm^{-1} and J refers to the lower state. Since $D \ll B$, the rotational absorption consists of a set of lines of almost equal spacing $2B$ (section 3.6). In diatomic molecules B gives directly the moment of inertia I and hence the bond length. In linear polyatomic molecules it is necessary to do experiments with different isotopes of one or more of the atoms to obtain the individual bond lengths unambiguously. The analysis of non-linear polyatomic molecules is naturally a good deal more complicated, but with suitable isotopic substitutions bond lengths and angles can usually be extracted to accuracies of about ± 0.005 Å (0.5 pm) and ± 30 arc minutes, respectively. This information is crucial for theories of molecular bonding.

9.3.2 HYPERFINE STRUCTURE AND QUADRUPOLE MOMENTS

As stated in section 3.3, most molecules have $^1\Sigma$ ground states, with no net angular momentum and consequently no magnetic hyperfine structure. The important hyperfine effect for molecules is the electric quadrupole interaction. There is usually an electric field gradient at the nuclei because of the cylindrical symmetry of the electron distribution. Taking the z-axis in the direction of this gradient and writing

$$q \equiv -\frac{\partial F_z}{\partial z}=\frac{\partial^2 V}{\partial z^2}$$

we have for the electric quadrupole interaction (equation 2.47)

$$\Delta E_Q = eQqf(F, I, J)$$

where eQ measures the departure from spherical symmetry of the nuclear charge distribution and $f(F, I, J)$ essentially describes the orientation of Q to the z-axis. The quantum numbers I and J determine the number of hyperfine levels (there is no splitting unless $I \geqslant 1$), and the quadrupole coupling constant eQq gives the magnitude of the splitting. If Q is known from other experiments, the microwave results can be used to evaluate q, an important parameter in molecular bonding theory. Even if Q is not known, useful information can be obtained from the relative values of q in different compounds containing the same nucleus.

9.3.3 STARK EFFECT

Most molecules show appreciable Stark splitting in an external electric field. As with the quadratic effect in atomic spectra (section 2.9) each level of given J splits into $J+1$ sub-levels, with $+M_J$ and $-M_J$ having the same energy. In linear molecules the splitting is proportional to p^2E^2, where $\boldsymbol{p}$ is the electric dipole moment and $\boldsymbol{E}$ the field strength. It is not linear in $\boldsymbol{p}\cdot\boldsymbol{E}$, as one might at first expect, because a uniformly rotating dipole has no average component in the field direction. The first effect of the interaction is to produce a net orientation in the field direction proportional to pE, and the interaction of the effective dipole $\boldsymbol{p}(pE)$ with the field $\boldsymbol{E}$ gives the quadratic dependence. p can therefore be determined from the Stark splitting. In a high field this may be as large as 20 MHz, measurable to about $1:10^3$, so the method is very accurate if E is accurately known. In practice the field is usually calibrated by measuring the splitting for a molecule of known p.

9.3.4 PARTICULAR CASES

The inversion spectrum of ammonia was the first microwave absorption studied, and its historical importance was enhanced when it became the basis of the first maser. The ammonia molecule, NH_3, is a pyramid with N at the peak of a base of H atoms, and the two configurations with the molecule turned inside out with respect to one another are both states of stable equilibrium. The resonance interaction between the two states leads to a splitting of about 24 000 MHz, corresponding to a wavelength of 8 mm.

The Lamb shift in hydrogen-like spectra also falls in the microwave region. The 2s–2p shift in H is about 0.03 cm^{-1}, or 1058 MHz (section 2.7). Before the era of Doppler-free laser spectroscopy it was impossible to measure such a small splitting in the H_α line with any degree of accuracy, but in 1947 Lamb and Retherford, in a classic microwave beam experiment, measured the energy difference directly with an accuracy of about $1:10^4$. Lamb shifts have since been measured for many other levels in the hydrogen-like ions.

9.4 Magnetic resonance spectroscopy: general considerations

The splitting of an atomic energy level into several sub-levels by interaction with a magnetic field, either external (Zeeman effect) or internal (hyperfine structure), is described in Chapter 2. As all such sub-levels belong to the same electronic configuration, transitions between them are magnetic dipole rather than electric dipole transitions.

To categorize the techniques involved, we start by looking at the orders of magnitude of the magnetic splittings. For a magnetic dipole μ in a field $\boldsymbol{B}$, the interaction energy is of the order of μB and the 'resonance frequency' of radiation required to induce transitions between sub-levels is given by $h\nu \simeq \mu B$. If μ represent the electronic magnetic moment, which is of order μ_B, and $\boldsymbol{B}$ is an external field of convenient size – say 0.1–1 T (1–10 kG) – the resonant frequency falls in the microwave region. This is the case for most free atoms. However, when atoms are bound into molecules or solids the electrons are usually paired off so that there is no resultant angular momentum and hence no electronic magnetic moment. Interaction with the external field then depends on the nuclear magnetic moment, some three orders of magnitude smaller; the interactions are correspondingly weaker, and the resonance frequencies are in the radiofrequency region. Finally, if the atom has both electronic angular momentum and a nuclear magnetic moment, the latter interacts with the *internal* magnetic field from the former; this is usually significantly stronger than any external field, so transitions between magnetic hyperfine levels tend to fall in the region of the microwave–radiofrequency borderline.

The amount of energy absorbed by a transition between two states decreases as the radiofrequency region is approached, and the measurement of absorbed energy therefore becomes an ever less sensitive method of detecting the transitions. In thermodynamic equilibrium the relative populations of two states separated by energy ΔE is given by Boltzmann's relation (equation 4.17):

$$\frac{N_2}{N_1} = \frac{g_2}{g_1} e^{-\Delta E/kT}$$

At room temperature $kT \simeq 200\ \text{cm}^{-1}$, and as the g-factors are close to unity the populations differ by less than 1% when ΔE is less than about $1\ \text{cm}^{-1}$. Consequently, stimulated emission is almost as probable as absorption. Spontaneous emission is negligible because of the ν^3 factor (equation 4.19). From equations 4.18 and 4.19 the net rate of upward transitions is

$$N_1 B_{12}\rho - N_2 B_{21}\rho = N_1 B_{12}\rho\left(1 - \frac{g_1}{g_2}\frac{N_2}{N_1}\right) = N_1 B_{12}\rho(1 - e^{-\Delta E/kT})$$

$$\simeq \frac{N_1 B_{12}\rho\, h\nu}{kT} \quad \text{for } h\nu \ll kT$$

Each upward transition means energy absorption of $h\nu$, so the rate of energy

absorption falls off as ν^2. Furthermore, the B-coefficient itself pertains to a magnetic dipole transition probability, which is smaller than an electric dipole transition probability by a factor of 10^4–10^5 (sections 2.8 and 11.3). Transitions for which the nuclear magnetic moment is responsible are down by a further factor of $(\mu_B/\mu_N)^2$, or about 10^6. Hence, energy absorption can be measured effectively only in solid or liquid samples where the density of absorbers is high. For free atoms and molecules a different method of detection is required, and the various atomic and molecular beam deflection, double resonance and level-crossing experiments provide far more sensitive methods of detection.

The information obtained from the magnetic resonance absorption type of experiment is of interest principally in solid-state physics, nuclear physics, chemistry and biology. We outline the principles of the two types of spectroscopy involved, electron spin resonance (ESR) and nuclear magnetic resonance (NMR), briefly in the next section before proceeding to the techniques relevant to free atoms and molecules.

9.5 Electron spin and nuclear magnetic resonance spectroscopy (ESR and NMR) [1, 2, 4]

In most solids and liquids the valence electrons are paired off to give zero resultant electronic magnetic moment, but there are two important types of exception: atoms or ions with partly filled inner shells (transition elements and rare earths) and free radicals *do* have unpaired bound electrons. Magnetic resonance spectroscopy with such substances is known as electron spin resonance (ESR) or, sometimes, paramagnetic resonance as it is these same unpaired electrons that are responsible for the paramagnetism. The reason for the name *spin* resonance is that the crystalline field is usually large enough to uncouple the orbital and spin angular momenta $\boldsymbol{L}$ and $\boldsymbol{S}$ and to separate the states of different L by several thousand wavenumbers in energy. The spin does not interact directly with the crystal field, and the $(2S+1)$ states for a given L are normally almost degenerate. However, the spin does interact with an external magnetic field with an interaction energy given to a first approximation by equation 2.37:

$$\Delta E_B = g_S \mu_B \boldsymbol{S}^* \cdot \boldsymbol{B} = g_S \mu_B B M_S \tag{9.1}$$

The magnetic dipole selection rule $\Delta M_J = \pm 1$ amounts in this case to $\Delta M_S = \pm 1$, so that

$$h\nu = g_S \mu_B B \simeq 2\mu_B B \tag{9.2}$$

With $B \sim 1$ T such transitions lie in the GHz region ($\lambda \sim 1$ cm). The principal interest of ESR lies in the departures from the approximate relation of equation 9.2. These come from residual interactions with the crystal field and therefore depend on the other ions, the crystal structure and the orientation of its axis

with respect to the B-field. Further information on interionic interactions is obtained from the widths of the resonances.

In the absence of unpaired electrons the dominant interaction with the external magnetic field is that of the nuclear magnetic moment μ_I – hence nuclear magnetic resonance (NMR). From equation 2.45

$$\Delta E_B = -\mu_I \cdot \boldsymbol{B} = -g_I \mu_N \boldsymbol{I}^* \cdot \boldsymbol{B} = -g_I \mu_N B M_I \tag{9.3}$$

and transitions between adjacent M_I levels are given by

$$h\nu = g_I \mu_N B \tag{9.4}$$

corresponding for realistic fields to frequencies in the MHz range. As one might expect, NMR is used for the measurement of nuclear g-values, but its applications are very much wider than this. The main interest is in the departures from the simple equation 9.4, due to partial screening of the nucleus by the bound electrons so that the nucleus does not see the full external B-field. These 'diamagnetic shifts' are determined by the chemical bonding, as also are the widths of the resonances. For atoms of known nuclear g-value the diamagnetic shifts can be interpreted in terms of the molecular structure, and there are obvious applications to biological molecules in which the simple atoms H, C, N and O predominate. NMR has become a powerful medical diagnostic tool.

Evidently both ESR and NMR can be used with substances of known g-value and shift to determine unknown magnetic fields – NMR for strong fields and ESR for weak fields. For example, the 'proton magnetometer', uses water or paraffin wax for precision measurements of B in terms of a frequency that can be extremely accurately calibrated.

9.6 Atomic and molecular beam magnetic resonance [1–4]

The detection of resonance absorption by direct measurement of absorbed energy is not, as we have seen, a sensitive technique in the radiofrequency (rf) region, especially for the intrinsically weak magnetic dipole transitions. The atomic and molecular beam deflection methods were developed to provide very much more sensitive detection, suitable for densities sufficiently low that the atoms or molecules can be regarded as 'free'. The information obtained from atomic beam experiments relates to the hyperfine structure of the free atoms and their interaction with an external magnetic field; for molecules the situation is rather more complicated because molecular rotation and diamagnetic effects have still to be taken into account. We shall describe the principles of the method for atomic beams and then point out the differences for molecular beams.

The technique exploits the fact that an atom with a magnetic moment passing through an inhomogeneous magnetic field is deflected in the direction of the field gradient by an amount that depends on the orientation of its magnetic

moment. Transitions affecting this orientation therefore vary the deflection. In this section we are concerned with transitions induced by an rf field at a resonant frequency corresponding to the energy difference between two magnetic sublevels in a region of homogeneous magnetic field. The basic experimental arrangement is shown in Fig. 9.2. The force in the z-direction acting on an atom with magnetic moment μ_J in an inhomogeneous field B is given by

$$F_z = \mu_z \frac{\partial B}{\partial z}$$

where μ_z is the component of μ_J in the field direction. In Fig. 9.2 the field in the three regions A, B and C is in the same direction. A and B are of equal length and have equal and opposite gradients, while C is a region of uniform field. An atom leaving the source in the direction shown is deflected up by the A-field so as to pass through the central slit and is deflected down an equal amount by the B-field, finishing up on the axis again. If μ_z is changed while the atom is in the C-field region, the B-field no longer brings it back to the detector (dotted line path), so there is a change in signal. Since the field *direction* (as distinct from its gradient) is the same in all three regions, the space quantization of μ_J does not change unless some other field is applied. This other field takes the form of an rf field in the C-region. For the simplest possible case, $J = \frac{1}{2}$, μ_z has the two possible values $\pm\mu_J$ as shown in Fig. 9.3. At the resonance frequency, given by

$$h\nu = 2\mu_J B_0 = 2(\tfrac{1}{2} g_J \mu_B) B_0 = g_J \mu_B B_0$$

(where B_0 is the C-field), an atom in either state may undergo a magnetic dipole transition to the other state and be lost from the beam when it gets into the B-region. Evidently transitions either way will contribute to the signal, in contrast with the techniques measuring energy absorption; all that matters is that the magnetic moment should be reversed. The experiment can, in fact, be run in two

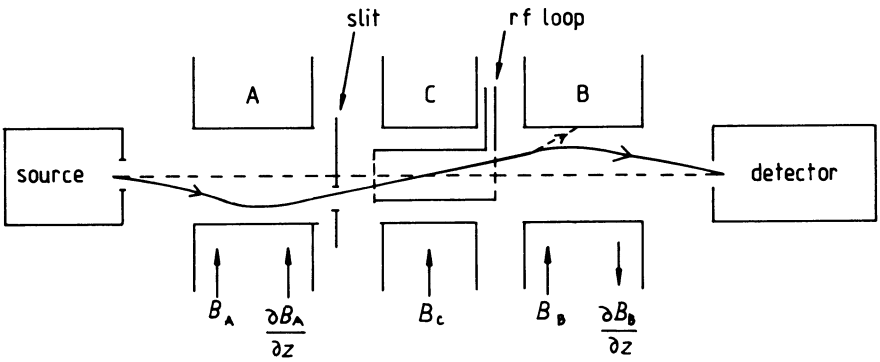

Figure 9.2 Schematic arrangement for atomic beam magnetic resonance. Distances perpendicular to the beam are greatly exaggerated. For explanation, see text.

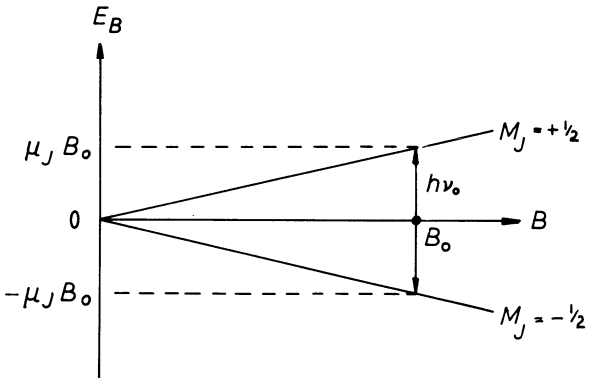

Figure 9.3 Resonance condition for an atom having $J = \frac{1}{2}$.

different modes: either 'flop-out' as just described or 'flop-in', with the gradients of the A- and B-fields in the same direction, so that resonance is detected by an increase in signal.

The source of atoms is usually an oven of some sort, producing atoms in the ground or a very low-lying excited state (< 1 eV), but in some cases more highly excited states have been produced by discharge techniques or laser excitation. Only metastable excited states are suitable because a lifetime of the order of a millisecond is required for the atom to get through the apparatus before it decays. The detector takes different forms: for a non-condensable gas it may be simply a pressure-measuring device; atoms such as the alkalis may be ionized at the surface of a hot filament and detected from the ionization current; other species may be ionized by electron bombardment and passed through a small mass spectrometer to sort out the beam from the background. Finally, radioactive atoms are detected simply by condensing them on a bit of foil and measuring the radioactivity.

The rf loop in the C-field is arranged to produce a magnetic component B_1 perpendicular to B_0, and it is B_1 that is responsible for the transitions. The fundamental width of the resonance is determined by the apparent duration of the radiation pulse – that is, by the time T that the atom spends in the rf field. According to the bandwidth theorem (or the uncertainty principle), the corresponding bandwidth is given by $\delta\nu \simeq 1/T$. For a velocity of about 10^3 m s^{-1} and a path of 0.1 m, $\delta\nu \simeq 10$ kHz. Increasing the path does not necessarily help because it is very difficult to keep B_0 sufficiently constant. This limitation can be overcome by using, instead of one long rf coil over the whole length of the C-region, a loop at each end, called Ramsey hairpins after the initiator of the method. If the total time of flight in the C-field is T and in each of the hairpins t, a sort of interference pattern results, with oscillations of width $1/T$ within an envelope of width $1/t$, as shown in Fig. 9.4. The effect is closely analogous to the magnification of aperture achieved by using two separated slits instead of one

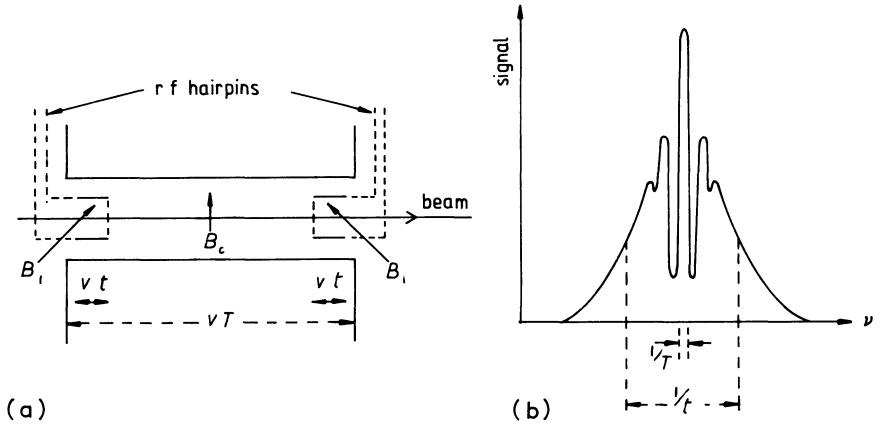

Figure 9.4 (a) 'Ramsey hairpin' arrangement of rf field, and (b) its effect on the signal. The broken parts of each hairpin are above the interaction plane, and B_1 is perpendicular to B_c. The signal has interference fringes of width $1/T$ superposed on the Lorentzian of width $1/t$.

large objective in the Michelson stellar interferometer. A similar trick, using laser instead of rf fields, is described briefly in section 8.6 as a method of achieving ultra-high resolution in optical spectroscopy.

Molecular beam deflection experiments differ from their atomic counterparts in that the magnetic moment responsible for the deflection is normally the nuclear moment, there being no resultant electronic angular momentum. Atoms with $^1S_0$ ground states are in the same category as molecules in this respect. The most important practical effect is that the deflections are smaller by a factor of $\sim 10^{-3}$ for a given field gradient. For $\partial B/\partial Z \sim 10^3\ \mathrm{T\,m^{-1}}$ ($10^5\ \mathrm{G\,cm^{-1}}$) over a path of 0.5 m, the deflection is only about 0.5 mm. Consequently, the whole apparatus has to be much longer, 4 m being fairly typical. The other differences following from the 1000:1 ratio of magnetic moments have already been remarked on in connection with NMR: the energy splittings are smaller and fall well into the rf region even in strong magnetic fields, and the transition probabilities are down by a factor of $(\mu_B/\mu_N)^2$, or about a million.

9.7 Transitions in and information from beam magnetic resonance [3, 5, 6]

Chapter 2 discussed the Zeeman effect (section 2.9) and hyperfine structure (hfs) (section 2.11) in atomic spectra, but did not deal with hfs in an external magnetic field, which is the case we are concerned with in this chapter. There are three possible interactions to be considered: the nuclear magnetic moment with the internal magnetic field B_J arising from the electron angular momentum (i.e.

field-free hfs), and the electron and nuclear magnetic moments separately with the external field. The relative importance of these interactions depends on the external field strength. Figure 9.5 shows the energy levels of an atom with unit nuclear spin ($I=1$) and a ground state having $J=\frac{1}{2}$ (lithium or deuterium, for example). In zero field μ_I interacts with B_J to give two hyperfine levels characterized by the quantum number F(where $\boldsymbol{F}=\boldsymbol{J}+\boldsymbol{I}$). In this case F can take two values, $\frac{1}{2}$ and $\frac{3}{2}$. Observed hyperfine splittings are in the range 0.01–1 cm^{-1}; since nuclear magnetic moments are of order 10^{-3} cm^{-1} per tesla, B_J must vary from about 10 T for light elements to 1000 T for heavy elements. So long as interactions with the external field are weaker than the hyperfine interaction, F remains a good quantum number, and the field splits each hyperfine level into $2F+1$ sub-levels, as shown in the centre of the figure. 'Weak' therefore implies $\mu_J B \ll \mu_I B_J$, or $B<0.01$ T for a light atom. The right-hand side of the figure shows the strong-field case, where $B \gg 10^{-3} B_J$ (0.1–10 T, say), and the μ_J–B interaction is dominant. The two groups of levels correspond to $M_J=\pm\frac{1}{2}$. However, so far as the nuclear moment is concerned B_J is still stronger than B, so μ_I interacts predominantly with B_J, and the three sub-levels $M_I=1, 0, -1$ may be thought of as hfs of the Zeeman effect. Only for $J=0$, the molecular

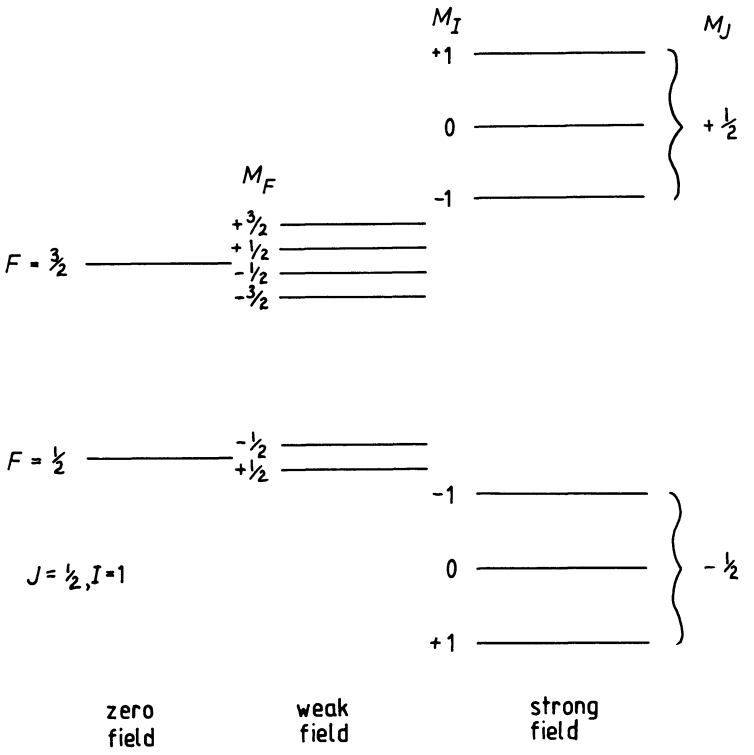

Figure 9.5 Hfs of atom with $J=\frac{1}{2}$, $I=1$ in zero, weak and strong magnetic field.

beam type of experiment, is the direct interaction of μ_I with B the first-order effect.

Figure 9.6 illustrates the hyperfine energy levels as a function of magnetic field for an atom with J again equal to $\frac{1}{2}$ but with an I-value of $\frac{3}{2}$. This is actually quite a general case because many of the one-valence-electron atoms with $J=\frac{1}{2}$ ground states have stable isotopes of spin $\frac{3}{2}$ (^{11}B, ^{23}Na, ^{39}K, ^{63}Cu, ^{69}Ga, ^{87}Rb, ^{197}Au, for example). From equation 2.46 the zero field hyperfine levels $F=1, 2$ are separated by $\Delta E_{IJ}=2A$, where the hfs constant A is proportional to B_J. In a weak external field each hyperfine level splits into $2F+1$ sub-levels defined by the magnetic quantum number M_F. The shift of each sub-level is given by an expression analogous to that for the ordinary Zeeman effect (equation 2.41):

$$\Delta E_B = g_F \mu_B \boldsymbol{F^*}\cdot\boldsymbol{B} = g_F \mu_B M_F B \tag{9.5}$$

The hfs g-factor, g_F, is a function of I, J and F corresponding to the Landé g-factor of equation 2.40. To a sufficient approximation

$$g_F = g_J \frac{F^{*2}+J^{*2}-I^{*2}}{2F^{*2}} \tag{9.6}$$

which for the particular case of $J=\frac{1}{2}$ simplifies to

$$g_F = +\frac{g_J}{2I+1} \quad \text{or} \quad -\frac{g_J}{2I+1}$$

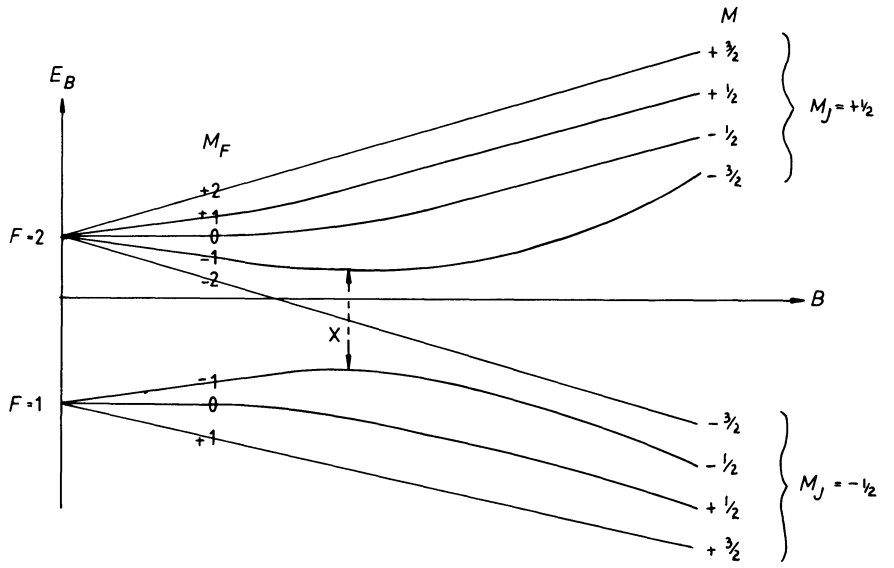

Figure 9.6 Hfs as a function of magnetic field for $J=\frac{1}{2}$, $I=\frac{3}{2}$. X marks the field-independent transition (see text).

for $F=I+\frac{1}{2}$ and $I-\frac{1}{2}$, respectively. Whatever the value of I, therefore, the spacing is the same in both sets of sub-levels, but the order of the M_F-values is reversed.

At the strong-field extreme the sum of the μ_J–B interaction and the weaker hyperfine interaction gives

$$\Delta E_B = g_J \mu_B B M_J + A M_I M_J \tag{9.7}$$

The hyperfine levels denoted by different values of M_I are almost equidistant and are independent of B, since we have ignored the smaller μ_I–B interaction. It can be seen from the figure that the magnetic quantum number of each sub-level, designated by M_F in the weak field and M_J+M_I in the strong field, remains constant.

The A- and B-fields are strong fields in this context, so that transitions, to be detectable, must involve a change of the M_J-value on the right-hand side of Fig. 9.6. The C-field is usually weak, and the transitions therefore have to satisfy the magnetic dipole selection rules $\Delta F=\pm 1$, $\Delta M_F=0, \pm 1$. All such possible transitions are shown in Fig. 9.7. The double lines in this figure are coincident when the spacing in the two sets of sub-levels is exactly equal. The broken lines correspond to $\Delta M_F=0$ (σ-components) and the full lines to $\Delta M_F=\pm 1$ (π-components). A cautionary note is necessary here: σ and π conventionally refer to the orientation of the *electric* vector of the rf field relative to the C-field, whereas magnetic dipole transitions are induced by the *magnetic* vector, which is

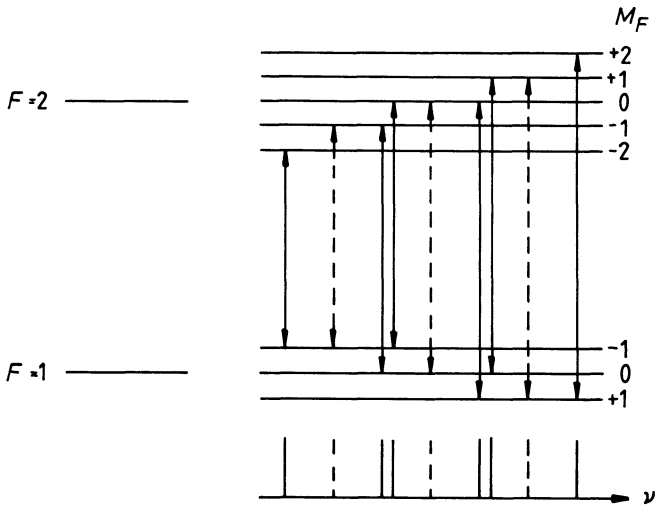

Figure 9.7 Weak field transitions for $J=\frac{1}{2}$, $I=\frac{3}{2}$. Transitions $\Delta M_F=\pm 1$ are denoted by solid lines and $\Delta M_F=0$ by broken lines. The double lines coincide when the spacing in the two sets of levels is equal, and all components are then equally spaced at intervals $h\nu=g_F\mu_B B$.

orthogonal to this. σ-transitions therefore require B_1 parallel to B_0; in the more usual case of B_1 perpendicular to B_0 only the π-transitions are induced (see Fig. 9.8 and accompanying explanation).

Measurements of these resonant frequencies can yield values for I, g_F, μ_I and the magnetic and electric quadrupole hyperfine coupling constants. In weak fields the value of I can be found either from the number of components or from their separation by way of g_F and equation 9.6. The centre of gravity of the pattern, which corresponds to $\Delta F = \pm 1$, $\Delta M_F = 0$, gives the zero-field hyperfine splitting, which can also be obtained by measuring the frequency of any one of the components as a function of B_0 and extrapolating back to zero field. For nuclei with $I > \frac{1}{2}$ and atomic states with $J > \frac{1}{2}$ electric quadrupole interactions will manifest themselves as small departures from equal spacing of the M_F levels;

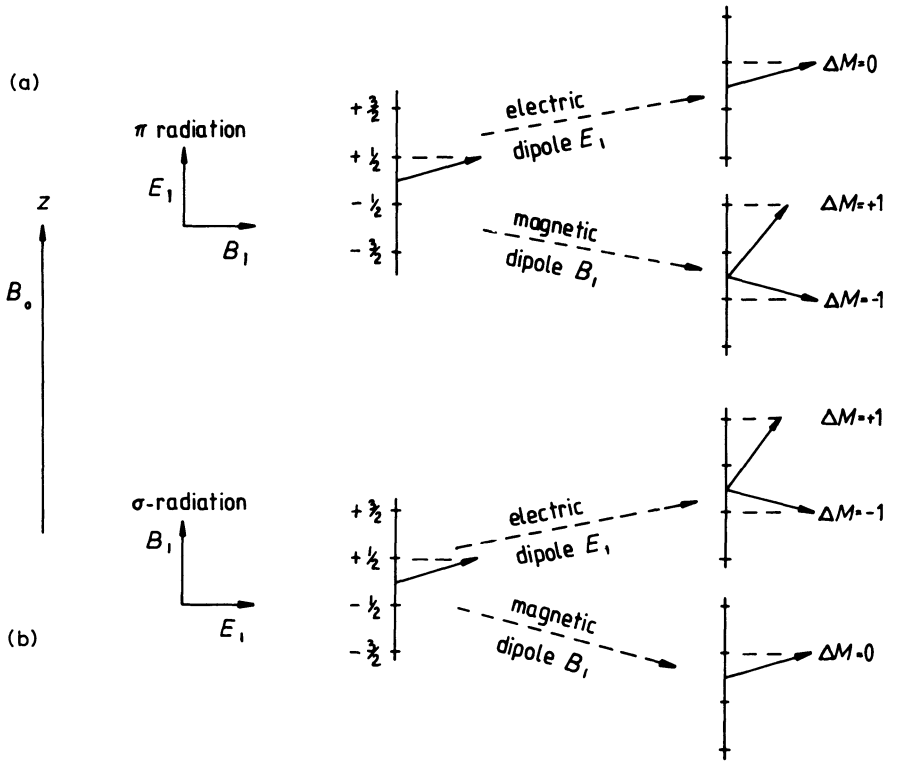

Figure 9.8 Rules determining the π- and σ-components for electric and magnetic dipole transitions. M can be changed only by the component of field perpendicular to B_0 – that is, by B_1 for π-radiation and by E_1 for σ-radiation. The diagram shows schematically the allowed transitions for an atom having J (or F) $= \frac{3}{2}$. The arrow gives the orientation of the vector $\boldsymbol{J}$ (or $\boldsymbol{F}$), starting in each case from the state $M = \frac{1}{2}$.

although it is eqQ that is measured (equation 2.47), the calculation of q necessary to evaluate the nuclear quadrupole moment eQ can be performed much more reliably for a free atom than for a molecule or solid (as required in the techniques described earlier in this chapter).

In strong C-fields the selection rules are $\Delta M_J = \pm 1$, $\Delta M_I = 0$, and if the M_I sub-levels were exactly equally spaced there would be four resonance frequencies spaced at equal intervals giving directly the value of the magnetic hyperfine constant A (see equation 9.7). The direct interaction between μ_I and the external field causes small departures from equal spacing, from which μ_I can be determined. The level-splitting and the resonance frequencies in intermediate fields can also be calculated and used to evaluate the same atomic and nuclear parameters. It is the results of such a calculation that are shown in Fig. 9.6; note that there is one particular transition, marked X on the figure, for which the resonance frequency passes through a minimum as a function of B. Such a 'field-independent transition'—and for $J > \frac{1}{2}$ there are usually more than one of them—allows particularly accurate measurements to be made because it is unaffected by small temporal and spatial variations in B.

Atomic beam magnetic resonance has found an important practical application in the atomic clock, now the established standard of time. The 'clock' is an atomic beam of ^{133}Cs, for which $I = \frac{9}{2}$ and J (ground state) $= \frac{1}{2}$. The resonance frequency for transitions between the $F = 3$ and $F = 4$ hyperfine states falls in a convenient part of the microwave region ($\lambda \simeq 3$ cm). The second is defined by taking this frequency to be 9192.631 770. . . MHz. The accuracy with which this can be determined is about one part in 10^{13}, corresponding to about 1 s in 200 000 years. This is several orders of magnitude better than astronomical time standards. The consequences for metrology, a fixed value for c and the abolition of an independent standard of length, are described in section 1.7.

Finally, the hfs of atomic hydrogen deserves special mention because it is the only case for which very accurate theoretical calculations can be made, using quantum electrodynamical theory. In the ground electronic state the two hyperfine levels $F = 1$ and $F = 0$ correspond to the electron and proton spins parallel and antiparallel, and the transition frequency is at 1420 MHz. For many years experiment and theory have leapfrogged over one another for the next significant figure in the measurement or the next term in the expansion. The transition is better known to astrophysicists as the 21 cm line radiated by interstellar hydrogen.

9.8 Double resonance and optical pumping experiments [7, 8]

In these techniques rf transitions between hyperfine or magnetic sub-levels of an excited state are detected from changes in the polarization (or absorption) of a resonance line in the optical region of the spectrum—that is, a line connecting the excited state to the ground state. The principle is most easily explained by a

specific example, for which we take the first double-resonance experiment ever performed, that by Brossel and Kastler on mercury. The ground state of the mercury atom is $6s^2\ {}^1S_0$, and as $J=0$ it has no hfs and no Zeeman splitting. The strong intercombination line at 253.7 nm connects the ground state to the excited $6s6p\ {}^3P_1$ level, for which $J=1$ and $M_J=+1, 0, -1$. In a magnetic field the line has three components as described in section 2.9: a π-component for $\Delta M_J=0$ and two σ-components for $\Delta M_J=\pm 1$. In the *double resonance* experiment mercury vapour in a magnetic field B_0 is illuminated with resonance radiation from a mercury lamp polarized in the B_0 direction as shown in Fig. 9.9. Since this counts as π-radiation for the vapour, it can excite only the $M_J=0$ sub-level of the P-state (Fig. 9.10). The radiation emitted when this decays back to the ground state necessarily has $\Delta M_J=0$ and is thus still polarized in the B_0 direction. A detector placed in this direction, in the position shown in Fig. 9.9, receives no signal because the direction of propagation of an electromagnetic wave must be perpendicular to the direction of polarization. Now if an rf field of the 'resonance frequency' given by $h\nu=g_J\mu_B B_0$ is applied, with its

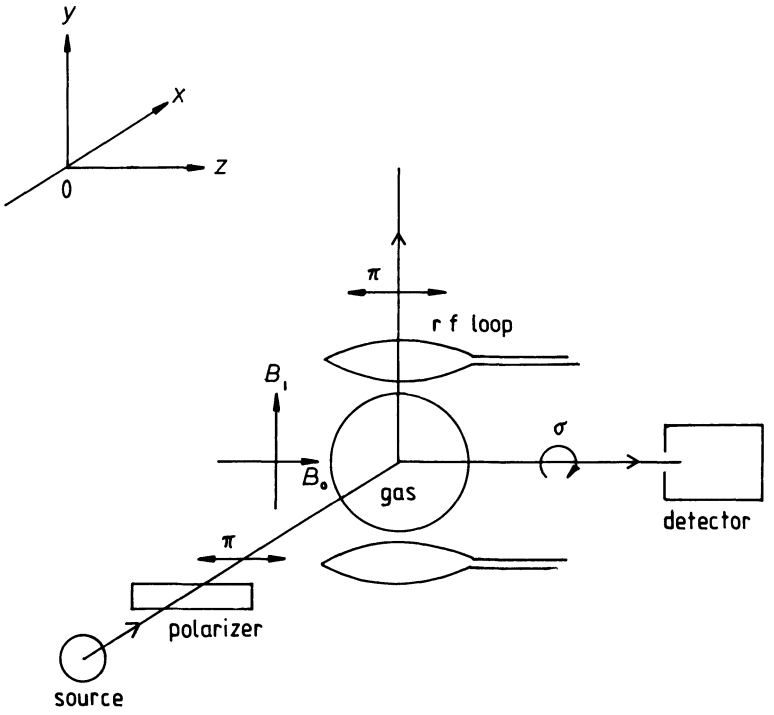

Figure 9.9 Schematic arrangement for a double resonance experiment. Resonance radiation polarized in the z-direction is incident along the x-axis. Re-emission in the z-direction is possible only if the rf field B_1 induces a transition (see Fig. 9.10).

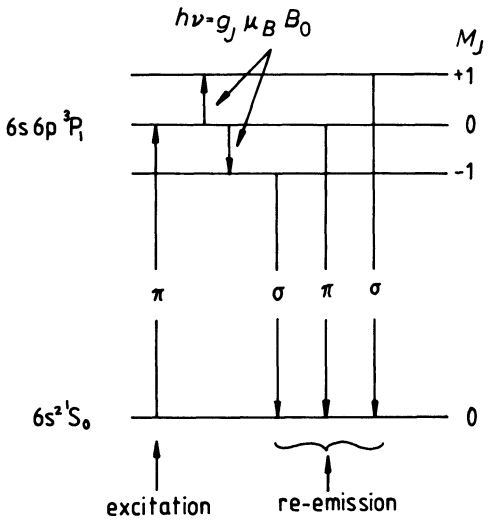

Figure 9.10 Transitions in double resonance experiment in mercury.

magnetic vector B_1 perpendicular to B_0, magnetic dipole transitions to $M_J = \pm 1$ are induced. The re-emitted 253.7 nm line then contains σ-components, as shown on the right of Fig. 9.10, which *can* travel in the direction of the detector. In this way the rf resonance is detected by the optical resonance – hence the name of the technique. Several variations of geometry and polarization are possible, but the essence of the method is the detection of a change in polarization.

This is an appropriate place to attempt to resolve confusion over the different selection rules governing π- and σ-components for electric dipole and magnetic dipole transitions, for the double resonance experiment is an example of both. By the usual convention, π means that the *electric* vector of the radiation is parallel to B_0 and σ that it is perpendicular, whatever the type of transition. In Fig. 9.8 B_0 is in the z-direction. If the rf field has E_1 in the z-direction and B_1 in the x- or y-direction, as in line (a), it counts as π-radiation. If B_1 is in the z-direction and E_1 in the x- or y-direction, as in line (b), it is σ-radiation. The selection rules shown schematically at the right of the figure follow from the fact that M is the component of magnetic moment in the B_0-direction and can be changed only by a vector perpendicular to that direction. Thus, in line (a) B_1 changes M and E_1 does not, and in line (b) the reverse is true.

The double resonance technique can be applied to the Zeeman sub-levels of hyperfine as well as fine structure. If the energy levels of Figs 9.6 and 9.7 are taken to belong to an excited state, rf transitions between them can alter the

polarization of the optical radiation to the ground state in exactly the same way as that described for Fig. 9.10. Magnetic dipole and electric quadrupole hyperfine interactions have both been measured by this method for atoms having structureless ($J=0$) ground states.

The intrinsic accuracy of any measurement of splitting is lower for an excited state than the ground state, partly because the population of an excited state is relatively small and partly because the sharpness of the resonance is limited by the natural lifetime of the excited state. For a lifetime of 10^{-8} s, typical of the upper level of a resonance line, the natural half-width is about 30 MHz, or 10^{-3} cm^{-1}. This is still a great improvement on the linewidths obtainable in optical spectroscopy except when Doppler-free methods can be used. The width may, in fact, be used to measure the lifetime of the excited state, provided that suitable corrections can be made for the power-broadening produced by high rf field intensity and for a phenomenon known as 'coherence narrowing'. The effect of the latter is to reduce the natural width when the gas pressure is high enough for more than one atom to interact with one photon.

The double resonance and atomic beam magnetic resonance methods have actually been combined in order to study excited states in atomic beams. A change of M_J while the atom is in the C-field can be achieved by exciting optically a higher state and inducing an rf transition in this state. The atom then decays to a different sub-level of the ground state and the resonance is detected by a change in deflection in the usual way.

The double resonance experiment just described is a particular case of a fairly large class of *optical pumping* experiments that exploit magnetic resonance without having recourse to atomic beam deflection methods. Resonance transitions between Zeeman and hyperfine sub-levels cannot be detected simply by monitoring the rf radiation absorbed because the populations, if thermal, are so nearly equal – the problem already enlarged upon in section 9.4. However, if the initial population is *non*-thermal, then changes can readily be detected. The preparation of a non-thermal population ratio is known as optical pumping, and an example is given in section 8.3, showing how the population of a $J=1$ ground state can be piled into the $M_J=+1$ sub-level (Fig. 8.4). It is not in the least essential to use a laser for optical pumping, and indeed all the early experiments were done with resonance radiation from conventional sources; but it is certainly very much easier to do them with tunable lasers.

The extent of the population imbalance can be measured either from the transparency of the vapour to light of the appropriate polarization (right-circularly polarized in the example of Fig. 8.4) or from the polarization of the fluorescence. The change in population induced by an rf field scanned through the resonance frequency shows up as a change in the absorption or the polarization. The effects of collisional and other relaxation processes must always be properly allowed for.

9.9 Level-crossing experiments [7, 8]

Level-crossing experiments have in common with magnetic resonance the investigation of fine or hyperfine sub-levels in a magnetic field, but they detect energy coincidences rather than transitions induced by an rf field. The method is most easily explained by a specific example.

Figure 9.11 shows hyperfine levels as a function of B for the case $J=1$, $I=\frac{1}{2}$. The rule of constant magnetic quantum number referred to in section 9.7, $M_F=M_J+M_I$, requires that two sub-levels change group in going from the weak- to the strong-field case, and in so doing they have to cross over. This diagram should be compared with Fig. 9.6, in which there are only two groups of sub-levels at both extremes; one sub-level changes group, and there is no crossing. For still larger values of J or I the number of sub-groups increases, and there is more than one crossing place. Now suppose we have an unsplit lower state, from which one sub-level of the pair shown in Fig. 9.11 can be selectively populated by appropriately polarized resonance radiation. For an arbitrary value of the magnetic field the re-emitted light has the same polarization but if B

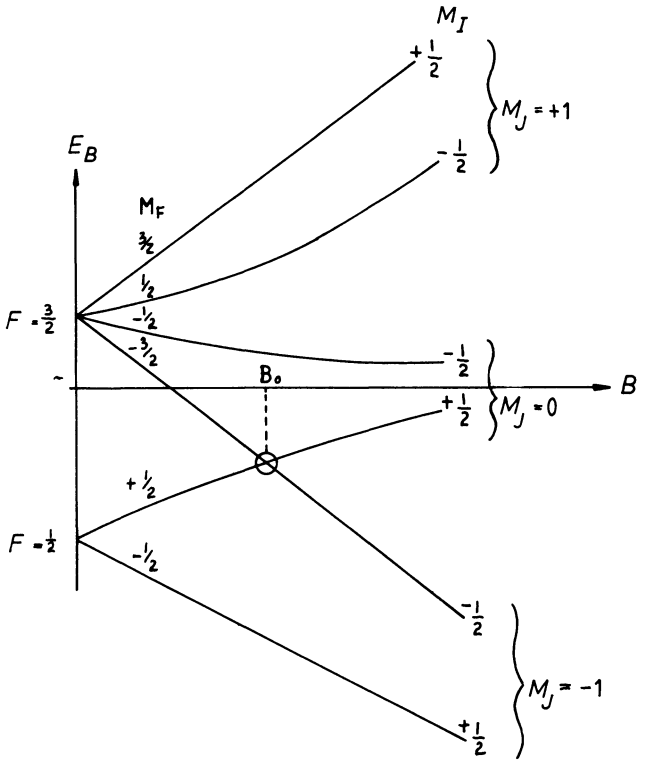

Figure 9.11 Level-crossing for $J=1$, $I=\frac{1}{2}$.

has the critical value B_0 the two sub-levels are degenerate. One can think of the emitted light as coming also from the second sub-level, and there is a partial change in polarization because of the different magnetic quantum number. Any splitting of the lower state must also be taken into account in calculating this change. In a proper quantum-mechanical treatment the upper state is treated as a superposition of two basis states which are degenerate at the particular field B_0 and radiate coherently; the appropriate transition matrix elements (cf. equation 2.33) then determine the change in polarization at each critical value of B.

If the polarization change is measured as a function of B, the width of the 'resonance' gives the lifetime τ of the excited state. This can be understood by attributing to each of the levels in Fig. 9.11 a finite width proportional to $1/\tau$. The two sub-levels overlap for a finite range of B each side of B_0, and τ can be found from the width of the range. This method of determining lifetimes is one of those in current use, although it is more often applied in the zero-field level-crossing technique otherwise known as Hanle effect, which is further considered in Chapter 12.

If the object of the experiment is to measure B_0 rather than the lifetime, the finite width may constitute a limitation on accuracy, particularly if there are several close components. It is possible to reduce the width by using a pulsed source for the exciting radiation and observing the re-emitted radiation after a finite delay T, where $T > \tau$. In effect the signal then comes only from those atoms that have survived longer than average in the excited state. This pulsed type of experiment can also give rise to so-called quantum beats: if the B-field is such that two sub-levels are very nearly degenerate, and if the exciting pulse populates them coherently, then there is an interference term in the wave function describing the upper state that results in a sinusoidal modulation of the radiated intensity. The frequency of the modulation gives directly the energy separation of the sub-levels. Both of these effects – delay-narrowing and quantum beats – were actually observed using conventional light sources for excitation and photon-counting methods to observe the extremely weak signals, but the very much larger fraction of atoms that can be initially excited with a pulsed tunable dye laser makes the experiments easier to perform and potentially much more accurate. The quantum-beat method is now one of the established ways of measuring small splittings in excited states.

9.10 Raman spectroscopy [9, 10, 12]

When atoms or molecules are irradiated by non-resonance radiation, a certain fraction of the radiation is scattered elastically – i.e. with wavelength unchanged. This process is known as Rayleigh scattering. The probability of its occurrence can be expressed in terms of a scattering cross-section which depends on the polarizability of the molecule and on the wavelength of the radiation. ('Molecules' in this context must be understood to include atoms, but it is molecules that are much more frequently involved in scattering processes.) Classically, one

can think of the bound electrons in the molecule as executing forced vibrations in the electric field $\boldsymbol{E}e^{i\omega t}$ of the electromagnetic wave with an amplitude proportional to E and to the polarizability, as a result of which they radiate perpendicular to their direction of motion at the frequency ω of the driving force. This same model is used in Chapter 11 to derive the expressions for refractive index and dispersion, which, indeed, result from coherent forward scattering. In a perfectly homogeneous medium the contributions to sideways-scattering cancel, and it is the statistical fluctuations about the mean density that are responsible for Rayleigh scattering. Its close connection with the refractive index n is brought out by the fact that the Rayleigh scattering cross-section is proportional to $(n-1)^2$.

There is a very much smaller probability for the molecules to scatter the radiation inelastically. The difference in energy of incident and scattered photons is made up by a change in excitation energy (rotational, vibrational or even electronic) of the molecule. This is Raman scattering, and the relation between the angular frequencies of the incident (ω) and scattered (ω_s) radiation

$$\hbar\omega - \hbar\omega_s = \Delta E \tag{9.8}$$

evidently gives information on the energy levels of the molecule. For a red shift, or ΔE positive, the molecule finishes in a higher state; for a blue shift, or ΔE negative, it finishes in a lower state. These two cases are known as Stokes and anti-Stokes scattering, respectively.

The classical picture of scattering can be extended to include Raman scattering. The polarizability of the molecule depends in general on its orientation relative to the electric field vector, and as the molecule rotates the polarizability, and hence the scattering cross-section, is modulated at twice the rotational frequency. The modulated scattered wave can be broken down into an unshifted component and two sidebands separated from it by the modulation frequency. A similar argument applies to vibration: the polarizability depends on the internuclear distance and is therefore modulated at the vibrational frequency, again giving rise to sidebands at a displacement equal to this frequency.

The classical picture is inadequate, because it accounts neither for the appearance of additional Raman components nor for the observed difference in intensity of the Stokes and anti-Stokes components. The quantum-mechanical model is the two-photon process shown schematically in Fig. 9.12. The molecule is excited by absorption of a photon $\hbar\omega$ to a virtual level. Insofar as this is energetically distant from a real level (large δE), the uncertainty principle, $\delta E \delta t \sim h$, allows the virtual state a vanishingly small lifetime; the molecule decays instantaneously by emission of a second photon either to its original state (Rayleigh scattering) or to a different real state (Raman scattering). For the Stokes lines the final states lie above the initial state, and for the anti-Stokes lines they lie below. The lower intensity of the anti-Stokes components is accounted for by the lower population of their initial (excited) states, as determined by the Boltzmann relation.

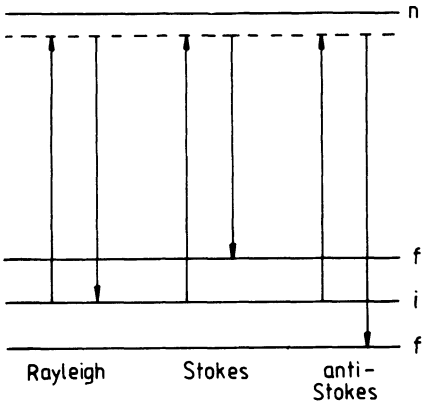

Figure 9.12 Rayleigh and Raman scattering as a two-photon process. The broken line represents a virtual level at some distance from a real level n.

In section 2.8 it is shown that first-order time-dependent perturbation theory gives a transition probability for emission or absorption of one photon in terms of the square of a dipole matrix element (equation 2.32):

$$P_{if} \propto E^2 |\langle f|\boldsymbol{er}|i\rangle|^2$$

where f and i represent the final and initial states. Two-photon scattering falls into the same category as the two-photon absorption process introduced in section 8.5. It must be treated by second-order time-dependent perturbation theory, and the scattering probability is given approximately by (compare equation 8.6)

$$P_{if} \propto \omega\omega_s^3 \left| \sum_n \frac{\langle f|\boldsymbol{er}|n\rangle\langle n|\boldsymbol{er}|i\rangle}{\omega_{ni} - \omega} \right|^2 \tag{9.9}$$

The transitions to and from the virtual state are, in effect, simulated by a transition from level i to a real intermediate stationary state n and a transition from n to f, summed over all intermediate states n. Equation 9.9 is a slight simplification, in that we have ignored polarization direction and omitted a second term that is normally small because it has $(\omega_{nf} + \omega)$ in the denominator. Clearly the sum is dominated by those real states nearest to the virtual state, and the probability rises rapidly when $\hbar\omega$ approaches a real state. This constitutes resonant scattering. P_{if} does not, of course, become infinite; as for two-photon absorption there is a damping term in the denominator that must be included near resonance.

The parity selection rule for Raman scattering can be deduced directly from equation 9.9: as each of the matrix elements requires a parity change, the net change must be zero. Thus, $\Delta l = 0, 2, \ldots$ for an electronic transition and $\Delta J = 0$ or ± 2 for a rotational transition. For a pure harmonic oscillator the selection

rule for v is $\Delta v = \pm 1$, as for a single-photon transition. The great importance of Raman scattering in molecular spectroscopy is that the matrix elements $\langle f|\boldsymbol{er}|n\rangle$, etc., do not vanish for homonuclear molecules, whereas the direct transition matrix element $\langle f|\boldsymbol{er}|i\rangle$ is zero for such molecules when f and i refer to the same electronic state (section 3.6). Those diatomic molecules which do not have pure rotational or vibrational spectra are therefore 'Raman active'.

In the case of pure rotation, $\Delta J = 0$ corresponds to elastic scattering, and a transition $J \leftrightarrow J+2$ gives by equation 3.16 (neglecting the D-term)

$$\sigma = B(J+2)(J+3) - BJ(J+1) = 2B(J+3)$$

where $J = J_i$ for the Stokes lines and $J = J_f$ for anti-Stokes. Figure 9.13 shows schematically a vibration–rotation Raman spectrum; the rotational line spacing is $4B$ as against $2B$ in the infra-red rotational spectrum. Polyatomic molecules also have pure rotational Raman spectra. Usually some of the vibrational modes are Raman-active and others give infra-red spectra.

The experimental problems of Raman spectroscopy are the low intensity of the inelastic scattering and the much greater intensity of the Rayleigh scattering. Raman cross-sections are typically $10^{-34}\ \mathrm{m^2\ sr^{-1}}$. For a gas at atmospheric pressure about 10^{12} incident photons are required for every inelastically scattered photon observed, whereas the Rayleigh scattering is perhaps five orders of magnitude greater. Raman spectroscopy was the first area to benefit from lasers; tunability is not necessary, and the first fixed-frequency lasers offered a huge advantage in photon flux over thermal light sources. The problem of looking for weak features close to the very strong Rayleigh line remains, of course, and it is usually necessary to use a double monochromator to discriminate against the background scatter. Obviously instruments of very large throughput are required, and 'Raman spectrometers' typically have input focal ratios of $f/4$ and rather low spectral resolution.

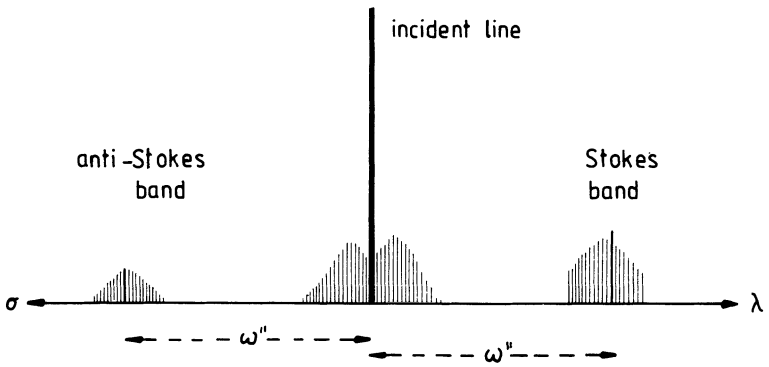

Figure 9.13 Schematic representation of Raman spectrum, showing rotational structure around the incident line and also in both a Stokes and an anti-Stokes band.

9.11 Coherent anti-Stokes Raman spectroscopy (CARS) [9, 11]

With the development of tunable lasers it has become possible to increase the efficiency of Raman scattering by using near-resonant incident radiation. The door has also been opened to more powerful techniques that exploit the coherence of the laser radiation and the non-linear properties of the scattering medium. The CARS technique described briefly in this section is the most used of these.

As seen in section 8.2, the non-linear terms in the polarizability of a medium are responsible for frequency mixing, so that sum-and-difference frequencies appear in the radiation propagated; in particular, the fundamental frequency can be doubled or tripled. CARS is a particular case of three-wave mixing, involving the third-order term in the polarizability. It is most easily envisaged as a four-photon process, as illustrated in Fig. 9.14. Three laser photons are required, two of which usually come from one laser at a fixed frequency ω_1. A second laser provides the third photon at a frequency ω_2 which can be tuned to satisfy the resonance condition

$$\omega_1 - \omega_2 = \frac{\Delta E_{12}}{\hbar} \tag{9.10}$$

The resonance is detected from the intensity of the signal at ω_0, where

$$\omega_0 = 2\omega_1 - \omega_2 \tag{9.11}$$

The appearance of this signal can be explained, in a somewhat oversimplified way, as follows.

If the electric dipole moment induced by an electric field $\boldsymbol{E}$ is expanded in terms of the higher orders of the polarizability, we have (equation 8.3)

$$\boldsymbol{p} = \alpha \boldsymbol{E} + \beta \boldsymbol{E}^2 + \gamma \boldsymbol{E}^3 + \cdots \tag{9.12}$$

α, β and γ are, properly speaking, tensors, but in an isotropic homogeneous medium such as a low-density gas they can be treated simply as coefficients, and

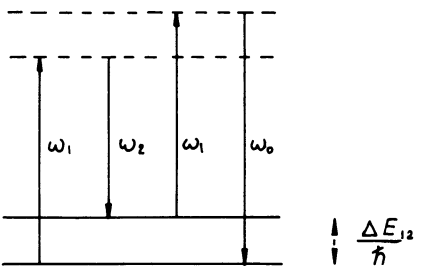

Figure 9.14 CARS as a four-photon process. ω_0 satisfies $\omega_0 = 2\omega_1 - \omega_2$, and the resonance condition is $\Delta E_{12} = \hbar(\omega_1 - \omega_2)$.

moreover β vanishes for symmetry reasons. Now if E is the sum of the electric fields from three coherent plane electromagnetic waves, $E_1 e^{i\omega_1 t}$, $E_2 e^{i\omega_2 t}$ and $E_3 e^{i\omega_3 t}$ – i.e.

$$E = \sum_{j=1}^{3} E_j \cos \omega_j t = \frac{1}{2} \sum_{j=1}^{3} E_j (e^{i\omega_j t} + e^{-i\omega_j t})$$

then E^3 contains all possible frequency combinations of the form $3\omega_1$, $2\omega_1 \pm \omega_2$, $\omega_1 \pm \omega_2 \pm \omega_3$, etc. The combination that is of principle interest for CARS is that given by equation 9.11. Picking just this one frequency, ω_0, the corresponding component of $\boldsymbol{p}$ becomes from equation 9.12

$$p(\omega_0) = \tfrac{1}{8}\gamma(\omega_0, \omega_1, \omega_2)\, E_1^2 E_2 \cos \omega_0 t$$

where the third-order polarizability γ is itself a function of ω_0, ω_1 and ω_2. By classical electromagnetic theory an oscillating dipole generates an electromagnetic wave at its oscillation frequency with power proportional to $\ddot{p}^2$. In this case the generated wave is a plane wave coherent with E_1 and E_2, with an $\boldsymbol{E}$-field given by

$$E(\omega_0) \propto \ddot{p}(\omega_0) \propto \omega_0^2 \gamma(\omega_0, \omega_1, \omega_2)\, E_1^2 E_2 \cos \omega_0 t$$

Assuming dispersion in the gas to be negligible, all three waves travel together and in phase through the gas. As each new molecule is set into oscillation it generates a field $E(\omega_0)$ coherent with the driving fields. A column of gas of unit area and length L containing N molecules per unit volume presents NL potential dipoles per unit area to the on-coming radiation, so by the end of the column the amplitude of $E(\omega_0)$ has been increased by the factor NL. As the flux density I (the power crossing unit area) is proportional to E^2, the flux density of the CARS signal is given in terms of those of the other two lasers by

$$I(\omega_0) \propto \omega_0^4 I_1^2 I_2 \gamma^2 N^2 L^2 \tag{9.14}$$

This signal can be as much as five orders of magnitude greater than that from spontaneous Raman scattering, and since it emerges as a laser-like beam with a divergence of a few mrad (instead of a spread of 4π sr) it can be more efficiently detected.

In spontaneous Raman spectroscopy it is the frequency of the scattered radiation, or rather its difference from the incident frequency, that provides the spectral information. In CARS the frequency of the generated signal is fixed by equation 9.11, and the spectral information comes from the resonant enhancement of the CARS signal in equation 9.14 whenever the variable frequency hits a value satisfying equation 9.10 for one of the energy differences in the molecule. The resonance condition is buried in γ. To disinter it requires a quantum-mechanical calculation that is well beyond the scope of this treatment, but it is possible to infer its existence by analogy with equation 9.9 for the two-photon case. As might be expected, γ turns out to be a sum of a very large number of

terms, each of which is the product of four matrix elements representing the four virtual transitions of Fig. 9.14 divided by the product of three frequency sums or differences. The resonance condition of equation 9.10 corresponds to a zero in one of these denominator terms. γ does not actually become infinite because, as usual, damping terms have to be included near resonance, but it does increase very significantly, greatly enhancing the intensity of the signal at ω_0. It can be shown that the magnitude of this signal at resonance is proportional to the square of the spontaneous Raman cross-section for the same pair of real levels, and also that all the selection rules are the same. The existence of all the non-resonant terms in the γ-sum is the principal disadvantage of CARS. They give rise to a non-resonant background signal which distorts the line shape and limits the ultimate sensitivity of the method.

CARS can evidently be used in the same way as spontaneous Raman spectroscopy to probe the structure of Raman-active molecules and, less frequently, atoms. The large signals make it possible to detect weak transitions, including those in excited states, and to work at high resolution with narrow-band CW lasers. However, it has a much wider range of applications, in that the high sensitivity obtained by using pulsed lasers of large peak intensity can be exploited to investigate gases at very low densities, transient species, or molecular fragments. Equation 9.14 shows that the CARS signal is proportional to the square of the column density of scattering molecules, providing a very sensitive test for the presence of trace species. The time resolution in pulsed mode allows rate processes to be studied.

References

These references have been grouped according to the topic principally covered, but there is considerable overlap between categories. Order within categories is alphabetical.

MICROWAVE, RADIOFREQUENCY AND MAGNETIC RESONANCE

1. Carrington, A. and McLachlan, A.D. (1979) *Introduction to Magnetic Resonance*, Chapman and Hall, London.
2. Ingram, D.J.E. (1967) *Spectroscopy at Radio and Microwave Frequencies*, Butterworths, London.
3. Ramsey, N.F. (1956) *Molecular Beams*, Oxford University Press (paperback edition 1985).
4. Walker, S. and Straw, H. (1961) *Spectroscopy*, Chapman and Hall, London.

MAGNETIC HYPERFINE STRUCTURE, etc.

5. Kopfermann, H. (1958) *Nuclear Moments*, Academic Press, New York.
6. Kuhn, H.G. (1969) *Atomic Spectra*, Longmans, London.

DOUBLE RESONANCE, etc.

7. Corney, A. (1977) *Atomic and Laser Spectroscopy*, Oxford University Press.
8. Series, G.W. (1981) 30 years of optical pumping, *Contemp. Phys.*, **22**, 487.

RAMAN AND CARS

9. Eesley, G.L. (1981) *Coherent Raman Spectroscopy*, Pergamon Press, Oxford.
10. Loudon, R. (1973) *The Quantum Theory of Light*, Oxford University Press.
11. Valentini, J.J. (1985) Coherent anti-Stokes Raman spectroscopy, in *Spectrometric Techniques IV* (ed. G.A. Vanasse), Academic Press, New York.
12. Weber, A. (ed.) (1979) *Raman Spectroscopy of Gases and Liquids*, Springer-Verlag, Berlin.

TEN

Width and shape of spectral lines

10.1 Line profiles

Any observed spectral line shape is a convolution of the true (source) line profile with the instrument function (Chapters 5–7) or the laser bandwidth (Chapter 8). Accurate measurements of wavelengths and intensities, as well as sensible choices of experimental parameters, require some knowledge of source line profiles. Moreover, these profiles are of interest in themselves for the information they can give on source conditions and for the study of interatomic interactions and collision processes. In this chapter we consider the true line shape, as produced by an ideal spectrometer of infinite resolving power.

Three different processes may contribute to the finite width of a spectral line: natural broadening, Doppler broadening and interactions with neighbouring particles. The last of these could take us all the way from low-pressure gases to the solid state, but the discussion here is limited to free atoms and molecules and comes under the description of pressure broadening. Even with this limitation the most that can be done in one chapter is to present a semiquantitative type of treatment.

Line shapes may be studied experimentally either in emission or absorption. If the line is scanned with an ideal spectrometer, the output signal I_ν will look something like Fig. 10.1, (a) in emission and (b) in absorption. The background continuum in the second case is assumed to have a constant value I_0 over the line profile.

The shapes of these curves are determined in the first instance by the frequency dependence of the emission and absorption coefficients of the source gas – j_ν and k_ν, respectively – but they can also be affected by the absorption and re-emission of radiation from the deeper layers on its way out. In this chapter we consider only the intrinsic line profiles represented by j_ν and k_ν, leaving the radiation transfer problem to Chapter 11. j_ν will reappear in section 11.9; it is defined as the power emitted per unit solid angle per Hz by unit volume of the gas. If all emitted radiation reaches the detector without reabsorption, the

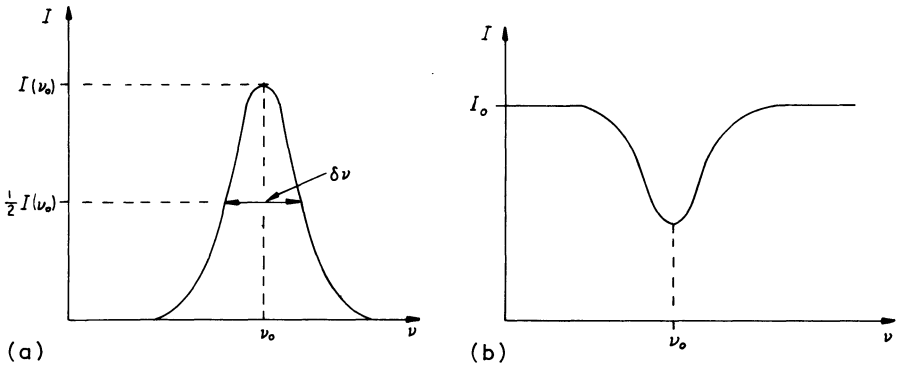

Figure 10.1 Line profiles as registered by an ideal spectrometer (a) in emission and (b) in absorption.

source is described as 'optically thin'. I_ν is then proportional to j_ν, and Fig. 10.1(a) may be regarded as a plot of j_ν against ν. The absorption coefficient k_ν is defined as the fractional decrease in flux density per unit path length:

$$-\mathrm{d}I_\nu = I_\nu k_\nu \,\mathrm{d}x$$

For a homogeneous layer of thickness l this integrates to give

$$I_\nu = I_\nu(0)\mathrm{e}^{-k_\nu l} = I_0 \mathrm{e}^{-k_\nu l} \tag{10.1}$$

k_ν is therefore found from Fig. 10.1(b) by calculating

$$k_\nu = \frac{1}{l}\ln\left(\frac{I_0}{I_\nu}\right) \tag{10.2}$$

at each frequency. This holds good whatever the depth of the absorbing layer. In the limit of an optically thin layer, for which $k_\nu l \ll 1$, equation 10.1 becomes

$$I_\nu = I_0(1 - k_\nu l)$$

or

$$k_\nu = \frac{1}{l}\frac{I_0 - I_\nu}{I_0} \tag{10.3}$$

so that the profile of k_ν is identical with the dip in Fig. 10.1(b).

The ratio j_ν/k_ν is known as the source function. For a given transition in a given environment it can normally be taken as constant over the line profile (Kirchhoff's law). Hence with an appropriate change of ordinate Fig. 10.1(a) may equally well be taken as a plot of k_ν against frequency.

The importance of any line-broadening process is usually measured by its full width at half the maximum (FWHM), shown as $\delta\nu$ in the figures (and represented by $\delta\lambda$ or $\delta\sigma$ if j and k are plotted as functions of wavelength or wavenumber). The line shape depends on the particular broadening process, and

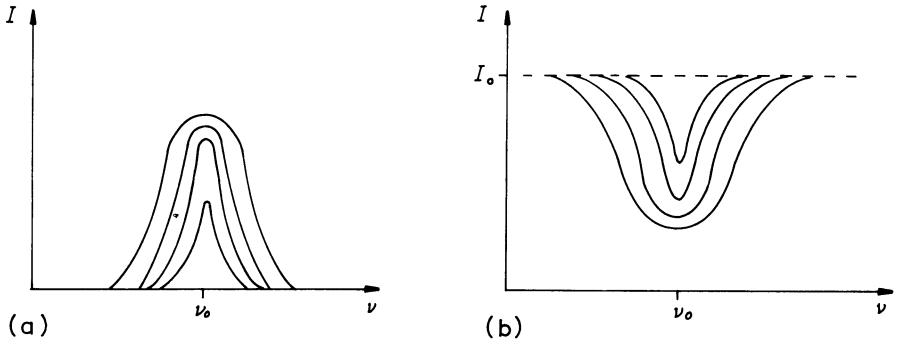

Figure 10.2 Effect of increasing optical depth on line profile (a) in emission and (b) in absorption.

the FWHM does not necessarily give useful information about the wings of the line. The term 'half-width' in the literature, incidentally, must be treated with some caution: while it usually has the meaning of *full* width at half-maximum, designated here as FWHM, it is occasionally taken to be half of this.

Figure 10.2 indicates how the profiles of emission and absorption lines change when the depth of the emitting or absorbing line is increased beyond the optically thin limit of $k_\nu l \ll 1$. In absorption it is obvious that as $k_\nu l \to \infty$ a saturation point is approached at which no light is transmitted at the central frequency ν_0. In emission the light from the centre of the line (where the absorption coefficient is highest) is preferentially absorbed on its way out, so the emission line peak also approaches a saturation value. These effects are treated more fully in section 11.9, but they are mentioned here because of their effect on the apparent width of the line. In investigating line-broadening processes in emission it is essential to use an optically thin source or to make appropriate corrections for self-absorption; and in absorption experiments k_ν must be found from equation 10.2 rather than equation 10.3 if there is any doubt about the optical depth.

10.2 Natural broadening

Natural broadening appears in both the classical and the quantum-mechanical theories of radiation. We look first at the quantum-mechanical picture in a semiquantitative way, but we derive the line shape classically.

Up to now the stationary states of atoms have been tacitly assumed to be infinitely sharp so that transitions between them have an infinitely narrow frequency spread. However, the uncertainty principle in the form $\Delta E \Delta t \sim \hbar$ requires an energy spread inversely proportional to the uncertainty in time associated with finding the atom in that particular state. Taking the mean

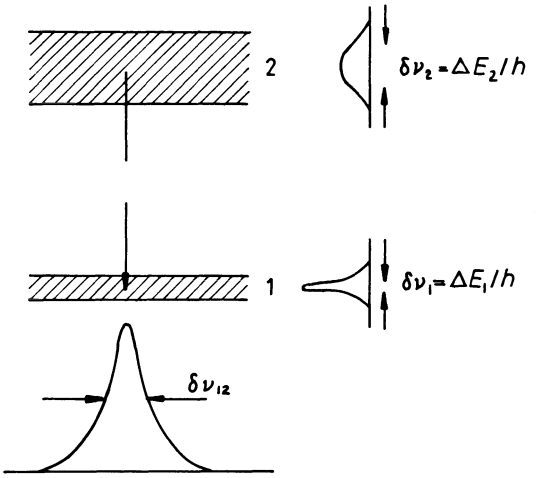

Figure 10.3 Natural broadening of spectral line. The energy levels are smeared out as indicated by the probability distributions on the right of the diagram. The FWHM of the line is given by $\delta\nu_{12} = \delta\nu_1 + \delta\nu_2$.

lifetime τ of the state as a measure of Δt, we have for the frequency spread of the state j

$$\delta\nu_j \simeq \frac{1}{2\pi\tau_j} \tag{10.4}$$

Whereas $\delta\nu_j$ is negligible for the ground or a metastable state ($\tau_j \to \infty$), the upper states of allowed optical transitions have lifetimes in the range 10^{-6}–10^{-9} s. Any spectral line starting or finishing on such a level must have a corresponding frequency spread, of order 0.1–100 MHz, or 10^{-5}–10^{-2} cm^{-1}. Figure 10.3 shows the situation when both levels are broadened. It is shown below that the width of the line is then given by

$$\delta\nu_{12} = \delta\nu_1 + \delta\nu_2$$

The lifetime of an excited state is the inverse of the spontaneous emission probability (assuming no collisional quenching): $\tau_2 = 1/A_{21}$. If more than one transition from level 2 is possible this expression is generalized to

$$\tau_2 = 1 \Big/ \sum_i A_{2i} \tag{10.5}$$

Owing to the ν^3 dependence of A_{21} (equation 4.19) natural linewidth decreases rapidly in the infra-red and microwave regions, but it may be appreciable in the far ultra-violet. In the visible region it is normally a couple of orders of magnitude smaller than Doppler broadening, but it is still possible for it to make a significant contribution to the wings of the line because the two effects have

different line shapes. The auto-ionizing levels mentioned in section 2.10 constitute an important exception to the rule of small natural width: the high probability of a radiationless transition makes for very short lifetimes, often of order 10^{-13} s, and correspondingly broad lines.

The shape of the broadened line depends on the shape of the energy distributions shown on the right-hand side of Fig. 10.3. These can be derived from Dirac's quantum theory of radiation, but in fact exactly the same line shape is obtained classically. As the classical oscillator appears again in Chapter 11 in connection with oscillator strength, absorption and dispersion, we introduce it here to derive the natural line shape.

In the classical picture the electron is a damped harmonic oscillator with a characteristic angular frequency ω_0 and damping constant γ. Its equation of motion is

$$\ddot{x}+\gamma\dot{x}+\omega_0^2 x=0$$

For small damping this has the well-known solution

$$x=a_0 e^{-(\gamma/2)t}\cos\omega_0 t=a(t)\cos\omega_0 t \tag{10.6}$$

The energy of the oscillator at time t is

$$U(t)=\tfrac{1}{2}m\omega_0^2 a^2(t)=U(0)e^{-\gamma t} \tag{10.7}$$

and the rate of loss of energy is

$$-\frac{\partial U}{\partial t}=\gamma U(0)e^{-\gamma t}=\gamma U(t) \tag{10.8}$$

γ can be evaluated by equating the energy loss rate to the power radiated by an accelerated charge according to classical electromagnetic theory:

$$-\frac{\partial U}{\partial t}=\frac{2}{3}\frac{e^2}{c^3(4\pi\varepsilon_0)}\overline{\ddot{x}^2} \tag{10.9}$$

Substituting for $\ddot{x}$ from equation 10.6,

$$\ddot{x}=-\omega_0^2 a(t)\cos\omega_0 t$$

and taking the cycle average we get

$$\overline{\ddot{x}^2}=\tfrac{1}{2}\omega_0^4 a^2(t)=\omega_0^2 U(t)/m \quad \text{by equation 10.7}$$

Therefore

$$-\frac{\partial U}{\partial t}=\frac{2}{3}\frac{e^2}{mc^3(4\pi\varepsilon_0)}\omega_0^2 U(t)$$

Equating these two expressions for $\partial U/\partial t$ gives for the damping constant

$$\gamma=\frac{2}{3}\frac{e^2\omega_0^2}{mc^3(4\pi\varepsilon_0)}=\frac{8\pi^2e^2\nu_0^2}{3mc^3(4\pi\varepsilon_0)}=\frac{2\pi e^2\nu_0^2}{3\varepsilon_0 mc^3} \tag{10.10}$$

The damped oscillations have the form shown in Fig. 10.4(a), with the amplitude decaying as $e^{-(\gamma/2)t}$. The corresponding energy decay is shown in Fig. 10.4(b) – an exponential with a time constant $1/\gamma$ which we define as τ_{class}. We can immediately deduce from the bandwidth theorem that a frequency spread of order $1/\tau_{class}$ should be associated with this time constant – i.e. $\delta\nu \simeq \gamma$, but we can actually do better than this. The amplitude $A(\omega)$ of the frequency spectrum corresponding to the time variation of x shown in Fig. 10.4(a) can be found by

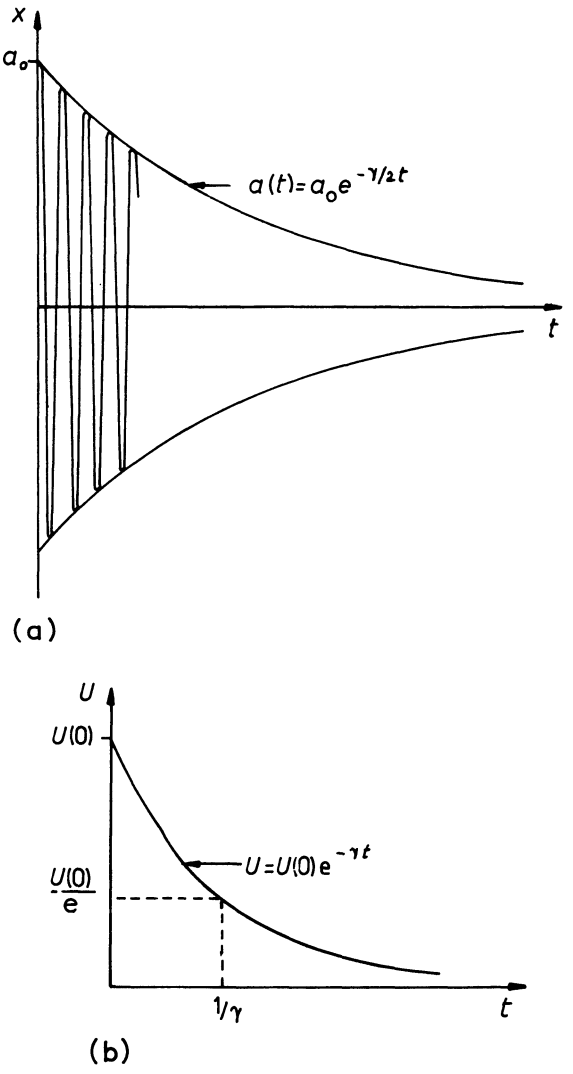

Figure 10.4 Damped classical oscillator showing decay of (a) amplitude and (b) energy with damping constant γ. The oscillation period in (a) is much exaggerated.

taking the complex Fourier transform of equation 10.6:

$$A(\omega)=\widetilde{x(t)}=a_0\int_{t=0}^{\infty} e^{-(\gamma/2)t}\cos\omega_0 t\, e^{i\omega t}dt$$

This integral can be readily evaluated by setting $\cos\omega_0 t=\frac{1}{2}(e^{i\omega_0 t}+e^{-i\omega_0 t})$. The intensity of the frequency spectrum is then found in the usual way by multiplying A by its complex conjugate. The result is

$$I(\omega)=A(\omega)A^*(\omega)\Rightarrow\frac{\text{const.}}{(\omega-\omega_0)^2+(\gamma/2)^2} \tag{10.11}$$

(There is actually a second term with $(\omega+\omega_0)^2$ in the denominator, which has been neglected because it is negligibly small by comparison.) This is the well-known Lorentzian or dispersion distribution. In terms of ν and the central intensity I_0 it becomes

$$I_\nu=I_0\frac{(\gamma/4\pi)^2}{(\nu-\nu_0)^2+(\gamma/4\pi)^2} \tag{10.12}$$

The shape of this distribution is shown in Fig. 10.5. The frequency $\nu_{1/2}$ at half intensity is given by

$$\nu_{1/2}-\nu_0=\pm\gamma/4\pi$$

leading to the FWHM

$$\delta\nu_N=\gamma/2\pi=\frac{1}{2\pi\tau_{class}}. \tag{10.13}$$

The full quantum-mechanical treatment also yields a Lorentzian line profile. If it is only level 2 that is broadened, the damping constant $\gamma=1/\tau_{class}$ is simply

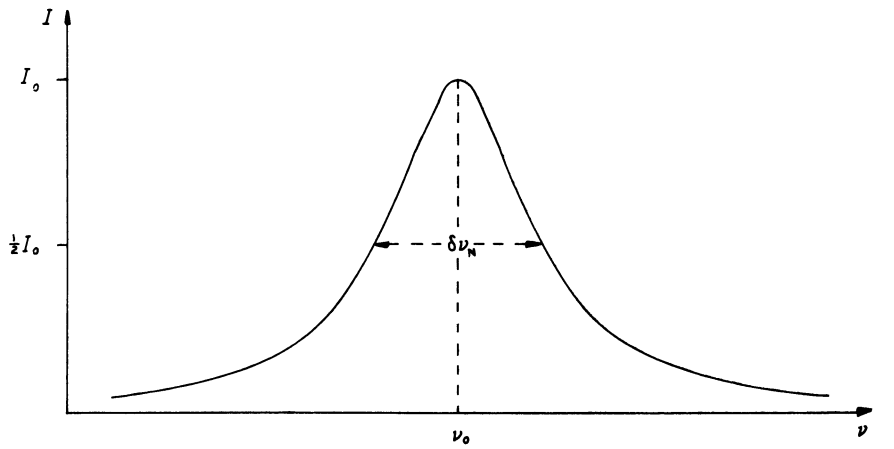

Figure 10.5 The Lorentzian, or dispersion, distribution, with FWHM $\delta\nu_N=\gamma/2\pi=1/2\pi\tau$.

replaced by $\gamma_2 = 1/\tau_2$, giving

$$\delta\nu_N = \frac{1}{2\pi\tau_2}$$

in agreement with the earlier estimate in equation 10.4. If level 1 also has a finite lifetime τ_1, its associated frequency spread $\gamma_1/2\pi$ contributes to the linewidth. The final line shape is obtained by convoluting the two contributions. This is one of the easier operations because the convolution of two Lorentzians is another Lorentzian with a damping constant given by $\gamma = \gamma_1 + \gamma_2$. The FWHM thus becomes

$$\delta\nu_N = \frac{1}{2\pi\tau_1} + \frac{1}{2\pi\tau_2}$$

Is there any relation between the classical lifetime and the quantum mechanical lifetime? Equation 10.10 implies that γ should depend only on the frequency, and in the visible region class works out to be about 1.6×10^{-8} s. Similarly, the classical natural width, if expressed in wavelength units, becomes a constant

$$\delta\lambda_{class} = \frac{e^2}{3\varepsilon_0 mc^2} = 0.016\,\text{pm}$$

Experimentally, lifetimes and natural widths are *not* constant. The linkage from $\delta\nu_N$ to τ_2 to A_{21} to B_{21} leads finally to the transition moments $\langle 2|er|1\rangle$ as the determining factors. $\delta\nu_N$ comes closest to the classical value for very strong lines having 'oscillator strengths' close to unity, as shown in Chapter 11. The transition between two states of an atom then corresponds closely with one classical oscillator. The resonance lines of the alkalis are good examples of such transitions.

10.3 Doppler broadening

Doppler broadening is a result of the well-known 'Doppler effect', which is the apparent shift in wavelength of the signal from a source moving towards or away from the observer – a decrease in wavelength for motion towards and an increase for motion away. The latter gives rise to the cosmological 'red shift' of the radiation from the receding galaxies.

In a laboratory source the shifts arise from the thermal motion of the emitting atoms or molecules. The observer sees a spread of shifts corresponding to the spread of velocities in the line of sight, and this is tantamount to broadening the line. We distinguished in section 8.3 between this 'inhomogeneous' broadening and the residual 'homogeneous broadening' exhibited by any subset of atoms with the same line-of-sight velocity.

The Doppler broadening line shape follows the velocity distribution of the atoms and can be derived easily for a gas in thermodynamic equilibrium. For an

emitter approaching the observer with a velocity v_x the Doppler shift is given by

$$-\frac{\Delta\lambda}{\lambda_0}=\frac{\Delta\nu}{\nu_0}=\frac{v_x}{c} \tag{10.14}$$

In equilibrium at temperature T the fraction of atoms with velocity between v_x and $v_x+\mathrm{d}v_x$ follows the Maxwell distribution

$$\frac{\mathrm{d}N(v_x)}{N}=\frac{\alpha}{\sqrt{\pi}}\mathrm{e}^{-v_x^2/\alpha^2}\,\mathrm{d}v_x \tag{10.15}$$

where α is the 'most probable velocity', given by

$$\alpha=\left(\frac{2kT}{m_A}\right)^{1/2}=\left(\frac{2RT}{M}\right)^{1/2}$$

In this expression m_A is the actual atomic or molecular mass, while M is the mass number; k and R are Boltzmann's constant and the universal gas constant, respectively.

Substituting v_x from equation 10.14 into equation 10.15, we have for the fraction of atoms emitting in the frequency interval ν to $\nu+\mathrm{d}\nu$

$$\frac{\mathrm{d}N(\nu)}{N}=\frac{1}{\alpha\sqrt{\pi}}\mathrm{e}^{-c^2(\Delta\nu)^2/\nu_0^2\alpha^2}\,\frac{c}{\nu_0}\,\mathrm{d}\nu$$

The intensity at ν is proportional to $\mathrm{d}N(\nu)$, so the line profile can be expressed in terms of the central intensity I_0 as

$$I_\nu=I_0\mathrm{e}^{-c^2(\nu_0-\nu)^2/\nu_0^2\alpha^2} \tag{10.16}$$

This is a Gaussian distribution about the central frequency ν_0 with a width determined by α. The half-intensity points are the values $\nu_{1/2}$ for which $I=I_0/2$ – i.e.

$$\frac{c^2}{\nu_0^2\alpha^2}(\nu_0-\nu_{1/2})^2=\ln 2$$

The FWHM is then

$$\delta\nu_\mathrm{D}=2|\nu_0-\nu_{1/2}|=2\sqrt{(\ln 2)}\,\frac{\nu_0\alpha}{c}$$

or

$$\delta\nu_\mathrm{D}=\frac{2\nu_0}{c}\sqrt{\left(\frac{2RT\ln 2}{M}\right)}=7.16\times10^{-7}\sqrt{\left(\frac{T}{M}\right)}\nu_0 \tag{10.17}$$

This can conveniently be expressed in the dimensionless form

$$\frac{\delta\nu_\mathrm{D}}{\nu_0}=\frac{\delta\lambda_\mathrm{D}}{\lambda_0}=\frac{\delta\sigma_\mathrm{D}}{\sigma_0}=7.16\times10^{-7}\sqrt{\left(\frac{T}{M}\right)} \tag{10.18}$$

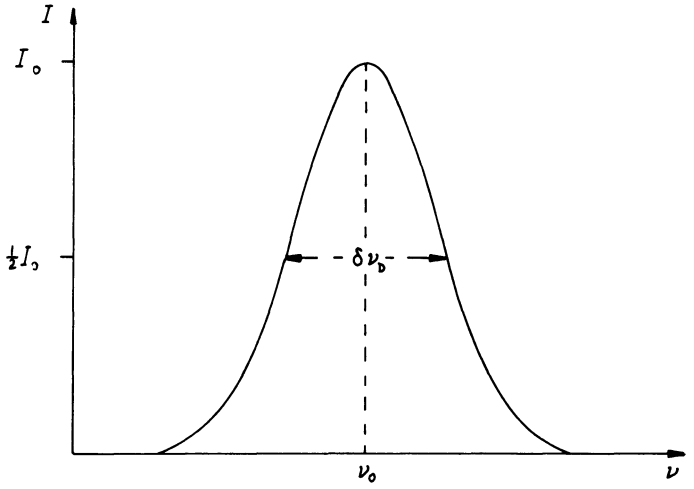

Figure 10.6 The Doppler, or Gaussian, distribution, with FWHM $\delta\nu_D = 7.16 \times 10^{-7}\sqrt{(T/M)}\nu_0$.

Equation 10.16 is also often put into the dimensionless form

$$I = I_0 \, e^{-y^2} \quad \text{with} \quad y = \frac{2(\nu_0 - \nu)}{\delta\nu_D}\sqrt{(\ln 2)} \tag{10.19}$$

The line profile is shown in Fig. 10.6. It falls off much more rapidly in the wings than does the Lorentzian profile, an important point to which we shall return later.

The range of Doppler widths can easily be found by putting specimen figures into equation 10.18. For atoms and diatomic molecules and for most laboratory sources M and T are, respectively, in the ranges 1–200 and 300–10 000, so the extreme values for the right-hand side of the equation are roughly 7×10^{-5} to 7×10^{-7}. In the visible region the corresponding widths are 0.04–0.0004 nm, or 0.1–0.001 cm^{-1}. The instrumental resolving powers required to match the Doppler widths are found from equation 10.18 simply by taking reciprocals, and with the above figures they range from 15 000 to 1.5×10^6. This is the basis for the assertion that an instrumental resolving power of more than a million is seldom useful. This limit does not, of course, apply to lasers, for which selection of one cavity mode within the Doppler profile results in a very much narrower linewidth.

In the past much effort has gone into reducing Doppler widths in order to obtain accurate peak wavelength measurements or to resolve close components or to study other broadening processes. For example, in low-current hollow cathode sources cooled with liquid nitrogen, or even liquid helium, the gas kinetic temperature can be well below 100 K, but the reduction in width is still

only a factor of about 3. An order-of-magnitude gain can be made by using a beam source, in which a series of collimated slits eliminates those atoms having an appreciable velocity in the line of sight, but the number density of atoms is then usually very low. Nowadays any problem requiring drastic reduction of Doppler width is likely to be tackled by one or other of the Doppler-free laser techniques described in section 8.6.

This discussion of Doppler broadening has assumed thermal kinetic motion, but a similar effect is produced by macroscopic motion in the form of turbulence or mass motions of the emitting gas. This effect is found in many astrophysical and some laboratory plasmas. In such cases the profile is not necessarily Gaussian.

10.4 Introduction to pressure broadening

It is a well-known experimental fact that spectral lines are broadened, and in many cases also shifted, by increasing gas pressure and by the presence of ions and electrons. Sometimes additional lines from 'forbidden' transitions also appear. These perturbations – widening, shifting and mixing of energy levels – are due to interactions between the emitting or absorbing atom or molecule and the other particles. The necessarily introductory treatment of this book can be followed up from the references at the end of the chapter, notably [5, 10] for neutral and charged particles, respectively. For an overview we follow the historical approach of treating two limiting cases of collisions: an almost instantaneous collision, on the one hand, and a quasi-static perturbation on the other. The 'unified theories', both semiclassical and quantum-mechanical, give entirely equivalent results at these two extremes. We shall therefore start by looking at the criteria for the two approximations and the extent to which they are met by different types of perturber (ion, electron or neutral particle).

The instantaneous collision extreme, known as the *impact approximation*, was first treated by Lorentz. He assumed that the collision cut off the wavetrain abruptly after a time t. From the usual lifetime probability arguments, the number of undisturbed atoms after a time t is related to the mean lifetime t_0 by the decaying exponential

$$N(t) = N_0 e^{-t/t_0}$$

This expression is identical with that describing spontaneous decay if t_0 replaces the lifetime τ_2 of the upper level, and the effect on the linewidth is exactly the same as in natural broadening: we obtain a Lorentzian distribution with a FWHM proportional to $1/t_0$.

The impact approximation is taken to the next stage of sophistication in section 10.6. The simple Lorentz treatment is a sufficient basis for consideration of its validity. It can only be expected to work if the collisions are discrete and well separated, so that the duration of a perturbation τ_c is small compared with

the time between collisions t_0 and each collision involves only one perturber. This last condition is known as the binary approximation. If the collisional width of the line is denoted by $\delta\nu_c$, we have as a measure of t_0 (by analogy with equation 10.4)

$$t_0 \simeq \frac{1}{2\pi\delta\nu_c}$$

and the required condition becomes

$$\tau_c < \frac{1}{2\pi\delta\nu_c} \tag{10.20}$$

For an estimate of τ_c consider Fig. 10.7, showing a perturber moving in a straight line with speed v past the radiating atom A. The distance of closest approach is defined by the impact parameter ρ. τ_c is conventionally measured by

$$\tau_c \simeq \rho/\bar{v} \tag{10.21}$$

where $\bar{v}$ is the mean speed. Writing equation 10.20 as

$$\delta\nu_c < \bar{v}/2\pi\rho$$

is not very helpful unless some upper limit is set on ρ. We therefore require ρ to be less than the value ρ_W at which the phase change produced by the perturbation is unity; this is taken to define a collision strong enough to destroy the coherence of the wavetrain (rather than actually terminating it as in the simple Lorentz treatment). The impact approximation validity condition then becomes

$$|\nu - \nu_0| < \bar{v}/2\pi\rho_W \tag{10.22}$$

with an additional condition to ensure that the atom interacts with only one perturber at a time

$$\tfrac{4}{3}\pi\rho_W^3 N < 1 \tag{10.23}$$

where N is the number density of perturbers. The reason for replacing $\delta\nu_c$ with $|\nu - \nu_0|$ in equation 10.22 is that the line wings are necessarily contributed by atoms that have had a lifetime which is shorter than average and may not satisfy the $\tau_c < t_0$ criterion. Equation 10.22 therefore becomes the criterion for judging how far from line centre one can expect the line shape to be the Lorentzian

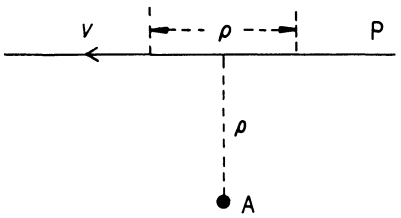

Figure 10.7 Impact parameter ρ for a perturber P moving with velocity v relative to a radiating atom A.

generated by the impact approximation. The conditions favouring it are:

(a) fast-moving perturbers→large $\bar{v}$;
(b) low perturber density→small N;
(c) short-range forces→small ρ_W.

From these considerations, the impact approximation may be expected to break down in the wings of the line first, for charged perturbers before neutrals (condition (c)), for ions before electrons (a) and at high densities (b).

The other extreme, first treated by Holtsmark, is known as the *quasi-static approximation.* The collision is now regarded as an almost constant perturbation over the time τ_c. Figure 10.8 shows the energy levels E_1 and E_2 of the radiating atom perturbed by the interactions $V_1(R)$ and $V_2(R)$, respectively, as a function of the perturber distance R. These curves can be thought of as the interatomic potentials of the quasi-molecule formed by the atom and its perturber, a system discussed in section 3.8.2. The similarity to Fig. 3.12(d) is evident. By the Franck–Condon principle, transitions take place at constant R, corresponding to vertical lines on the diagram. The transition frequency is shifted from the unperturbed value ν_0 by the amount

$$\Delta\nu(R)=\nu_0-\nu(R)=\frac{1}{h}[V_2(R)-V_1(R)] \tag{10.24}$$

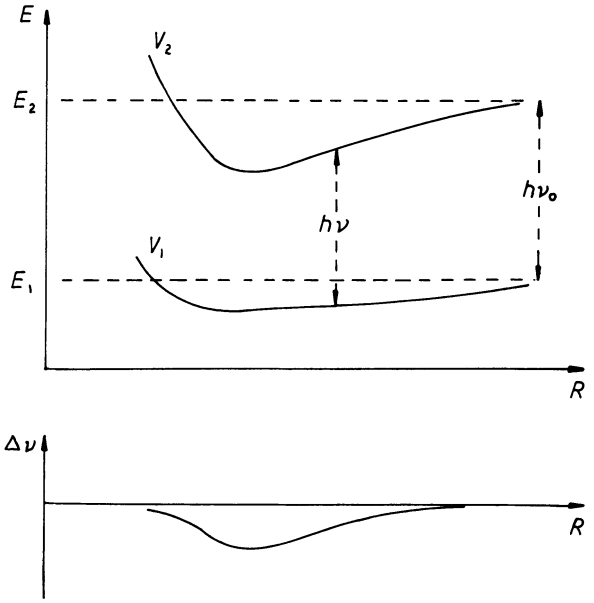

Figure 10.8 Perturbations V_1 and V_2 of the energy levels E_1 and E_2 as a function of perturber distance R. The lower diagram shows the resulting frequency shift $\Delta\nu=(V_2-V_1)/h$.

The line broadening is found by averaging the shifts $\Delta\nu(R)$ over the statistical distribution of atom–perturber separations R. Usually the higher level is more sensitive to perturbation, and one can think in terms of the change in energy of this level only:

$$\Delta\nu(R) \simeq \frac{1}{h} V_2(R)$$

The frequency shift from any given perturber reaches its maximum, stationary, value at the distance of closest approach, $R = \rho$. The validity of the quasi-static approximation depends on each atom being allowed to radiate at its shifted frequency $\nu(\rho)$ for long enough that the broadening associated with the finite collision time τ_c is small compared with $\Delta\nu(\rho)$ i.e.

$$\Delta\nu(\rho) > \frac{1}{2\pi\tau_c} \tag{10.25}$$

This condition, $2\pi\Delta\nu\tau_c > 1$, implies a phase change greater than unity, which, by the definition of ρ_w, means $\rho \leqslant \rho_w$. Thus the validity condition for the quasi-static approximation becomes

$$|\nu - \nu_0| \geqslant \Delta\nu(\rho_w) \quad \text{where} \quad \Delta\nu(\rho_w) > \frac{\bar{v}}{2\pi\rho_w} \tag{10.26}$$

which is just the opposite of the condition for the impact approximation. Obviously equation 10.26 is more likely to hold in the wings of the line and is favoured by low perturber velocity and long range forces ($\rightarrow$ large ρ_w).

The assumption of single perturbers simplifies the calculations but is not essential. Multiple interactions become increasingly important as the range of the forces increases.

Before going on to discuss these ideas in more detail, it is worth making a comment about *overlapping lines* that applies to any form of pressure broadening. Lines that are sufficiently pressure-broadened to overlap can no longer be treated as independent. In particular, the intensity in the overlap region is not just the sum of the two independent intensities. The reason is that the wave functions of the energy levels concerned are mixed by the perturbation, and it is necessary to choose a new basis set to evaluate the transition. Alternatively, one can think of the two wave functions as becoming partly coherent so that they can interfere with each other.

10.5 Pressure broadening: types of perturbation

An interaction which is a function of R and tends to zero for large R may be expressed as a power series in $1/R$:

$$V(R) = \sum_k C_k R^{-k} \tag{10.27}$$

The type of interaction determines the first non-vanishing term in this series,

$$V(R)=\frac{C_n}{R^n}+\text{terms of shorter range} \tag{10.28}$$

and the value of R for which $V(R)$ is significant sets the range of the interaction. Three different types of perturber may be distinguished.

(a) Charged particles

An emitting atom at a distance R from an ion or electron is perturbed by an electric field $E=e/(4\pi\varepsilon_0 R^2)$, and the interaction between the atom and the field is described by the Stark effect. For this reason broadening by charged particles is often described as Stark broadening. It is shown in section 2.9 that a perturbation proportional to E exists only in the case of the hydrogen atom; for all other atoms the first non-vanishing interaction is the quadratic Stark effect, proportional to E^2 and thus to $1/R^4$. For hydrogen and the H-like ions we therefore have $n=2$, with a constant C_2 determined by the linear Stark coefficients, which can be calculated from the hydrogen wave functions. For all other atoms we have $n=4$; the constant C_4 can in principle be calculated from the relevant wave functions, but is more usually found from the experimental determination of the quadratic Stark coefficients in a static external electric field.

The linear Stark effect splits the energy levels symmetrically, giving a symmetric line pattern. The statistical averaging process smears out this line pattern to give a symmetrically broadened, unshifted line. Conversely, the quadratic effect splits the levels asymmetrically and also shifts their centres of gravity, usually downwards (Fig. 2.17). The shift is normally greater for the higher states, so lines broadened by the quadratic effect are usually shifted to longer wavelengths and are often asymmetric. The direction of the shift is the same for ions as for electrons because of the E^2 dependence.

(b) Resonance

Resonance interactions occur only between identical atoms, and are effectively confined to states combining with the ground state. They take the form of a dipole–dipole interaction, for which n in equation 10.28 is 3. An atom in a stationary state has no permanent dipole moment, so this form of interaction needs some explanation. Consider the quasi-molecule A*B formed by an excited atom A* and a nearby identical atom B in the ground state. Because of the identity of the atoms, this system is degenerate with AB*, and the correct non-degenerate wave functions are

$$U_\pm=\frac{1}{\sqrt{2}}(u_{\mathrm{A}^*}u_{\mathrm{B}}\pm u_{\mathrm{A}}u_{\mathrm{B}^*})$$

implying that both A and B are partly in the ground state and partly excited. (There is a close analogy to the sharing of excitation energy between the two

electrons in a helium atom and to the sharing of an electron between the two protons in a H_2 molecule.) The expectation value of the electric dipole moment of the mixed state does not vanish; indeed its square, $|\langle U_{\pm}^{*}|\boldsymbol{er}|U_{\pm}\rangle|^2$, is proportional to the transition probability and hence to the oscillator strength f of the transition (see section 11.5 for the definition of f). The perturbation C_3/R^3 can be envisaged as an interaction between two dipoles each with a magnitude proportional to $\sqrt{f}$. C_3 is then proportional to f and depends on the orientation of the dipoles with respect to the interatomic axis – in other words, the m-values. Positive and negative values are equally likely, so resonance interactions give symmetrically broadened, unshifted lines.

(c) Neutral particles

The interaction between two neutral atoms takes us straight into the realms of interatomic potentials mentioned in Chapter 3. Each potential is peculiar to the particular pair of atoms involved; but, just as the Morse potential gave a useful approximation for the calculation of vibrational energies, so the van der Waals potential will serve for a semiquantitative discussion of the 'unbound molecules' that we are dealing with here. The van der Waals force actually acts between any two atoms or molecules, but it has a smaller range than either of the other two types and can be neglected in cases (a) and (b). It takes the form of an interaction between a fluctuating dipole in the radiating atom and the dipole that this induces in the perturber. Although the atom in a stationary state does not have a permanent dipole, its root mean square dipole moment is not zero – i.e.

$$\langle \boldsymbol{er}\rangle = 0$$

$$(\langle e^2r^2\rangle)^{1/2} = \sqrt{\overline{p^2}} \neq 0$$

The magnitude of the electric field E' that this produces at a neighbouring atom is given (within a small numerical factor) by

$$E' = \frac{\sqrt{\overline{p^2}}}{(4\pi\varepsilon_0)R^3}$$

If α is the polarizability of this second atom (the perturber), the induced dipole is $\alpha \boldsymbol{E}'$. The interaction between this induced dipole and the field is

$$V = -\alpha \boldsymbol{E}' \cdot \boldsymbol{E}' = -\frac{\alpha\overline{p^2}}{(4\pi\varepsilon_0)^2}\frac{1}{R^6}$$

This is always negative, corresponding to an attractive force. It is usually greatest for the most highly excited levels, so Fig. 10.8 can again serve as an illustration, and the perturbation corresponds to a red shift.

Certain perturbers, notably helium, have such small polarizability that the attractive force is almost zero and the first non-vanishing term in equation 10.27 is repulsive, corresponding to the steeply rising parts of the potential curves in Fig. 10.8. The perturbation is then a blue shift. Obviously repulsive forces come

into play in all types of perturbation at small enough R when the electron clouds begin to overlap, but they are negligible at the larger values of R relevant to cases (a) and (b). In case (c) they are often taken into account by using the Lennard–Jones potential

$$V=\frac{C_6}{R^6}+\frac{C_{12}}{R^{12}}$$

where $C_6<0$ and $C_{12}>0$. The choice of $n=12$ is for mathematical convenience rather than physical significance. Calculations prove to be fairly insensitive to the exact power chosen, and in any case this term can usually be ignored for values of R greater than two atomic radii, say 0.3–0.5 nm. For comparison, the average distance between atoms or molecules in a gas at STP is about 3 nm.

(d) Orders of magnitude

It is useful at this stage to look at the ranges of the interactions just discussed and at the validity of the binary approximation in each case. For this purpose we anticipate the expression for the Weisskopf radius given by equation 10.31:

$$\rho_W=\left(\frac{2\pi C_n}{h\bar{v}}\right)^{1/(n-1)}$$

The value of $\bar{v}$ to be used depends principally on whether the perturbers are electrons or neutral atoms and ions. It will suffice to take three 'typical' cases. For electrons at temperatures $T\sim 10\,000$ K, $\bar{v}$ is of order 6×10^5 ms^{-1}. For moderately heavy atoms or ions ($M\sim 60$) $\bar{v}\sim 2000$ m s^{-1} at 10 000 K and $v\sim 300$ m s^{-1} at room temperature.

Starting with the shortest range, case (c), we have

$$C_6\simeq\frac{\alpha\overline{p^2}}{(4\pi\varepsilon_0)^2}$$

The atomic polarizability is related to the refractive index by $n-1=N\alpha/2\varepsilon_0$ (a relation that can be deduced from section 11.5), and in the visible region $(n-1)$ is of order 3×10^{-3}. As an order of magnitude for p we can take the product of e and a typical atomic radius for a moderately excited state, say $3a_0$. These numbers make $C_6\simeq 10^{-77}$ J m^6. The appropriate $\bar{v}$ is that for neutral atoms at moderate temperatures. ρ_W is then about 0.5 nm. Experimental values of C_6 are actually in the range 10^{-77}–10^{-79} J m^6, and because of the 6th-power dependence ρ_W is very insensitive to the exact value, or to the value of $\bar{v}$ chosen.

The next-shortest range interaction is the quadratic Stark effect, for which we expect an interaction proportional to the polarizability of the atom and the square of the field strength:

$$V\simeq\frac{\alpha e^2}{(4\pi\varepsilon_0)^2R^4}\quad\text{or}\quad C_4\simeq\frac{\alpha e^2}{(4\pi\varepsilon_0)^2}\sim 5\times 10^{-58}\text{ J m}^4$$

In this case we are concerned only with electrons and ions at plasma temperatures. For ions ρ_W comes out to be about 1 nm and for electrons about 0.15 nm.

For the resonance interaction we expect something of the order of a dipole–dipole energy, $p^2/(4\pi\varepsilon_0 R^3)$. Putting $p \simeq ea_0$, we get $C_3 \sim 5 \times 10^{-49}$ and ρ_W around 2–4 nm, depending on temperature. The linear Stark effect has a still longer range, and one can often take for the cut-off of this interaction the Debye radius, as discussed in section 13.2.

For the binary approximation to hold there must not be more than one particle within a sphere of radius ρ_W. This condition is met for the van der Waals interaction at STP (mean interparticle distance $\simeq 3$ nm), but lower pressures of the gas or vapour under investigation are likely to be needed in the case of resonance broadening. ρ_W for electrons is so small that the binary approximation is likely to be valid in almost all conditions. In the case of ions the combination of large ρ_W with small $\bar{v}$ often justifies the use of the quasi-static approximation throughout the line profile, for in the case of the linear Stark effect, and the problem becomes one of dealing with multiple stationary perturbers rather than multiple collisions. In any event, an escape clause from the binary condition is available for the far wings of the line because these are attributable to very close perturbers and are little affected by more distant interactions.

10.6 Pressure broadening: the impact approximation

Current versions of impact broadening theory are extensions of Weisskopf's model of phase-changing collisions. Three basic assumptions are made, known as the binary, classical path and adiabatic approximations. The first of these has already been discussed. The second assumes the perturber to move along a classical path, which is a straight line unless both perturber and emitter are charged. The third implies that the perturbers do not induce transitions between, or mix up, different close-lying states of the emitting atom. Collisions are assumed to cause phase changes in the emitted wavetrain, but not to cut it short by knocking the radiating atom out of its excited state. One would expect that both processes might actually occur, but it turns out that for neutral perturbers the adiabatic approximation is usually good. When electrons are present, inelastic collisions may contribute appreciably to the linewidth, as will be seen later.

The frequency change $\Delta\nu(R)$ shown in Fig. 10.8 is responsible for the phase change in the wavetrain. As R decreases to its minimum value of ρ (the impact parameter) and then increases again, the frequency changes as shown in Fig. 10.9. The total phase change over the duration of the collision depends on ρ and is given by

$$\eta(\rho) = \int_0^{\tau_c} 2\pi\Delta\nu(R)\,\mathrm{d}t = 2\pi \int_{-\infty}^{\infty} \Delta\nu(R)\,\mathrm{d}t$$

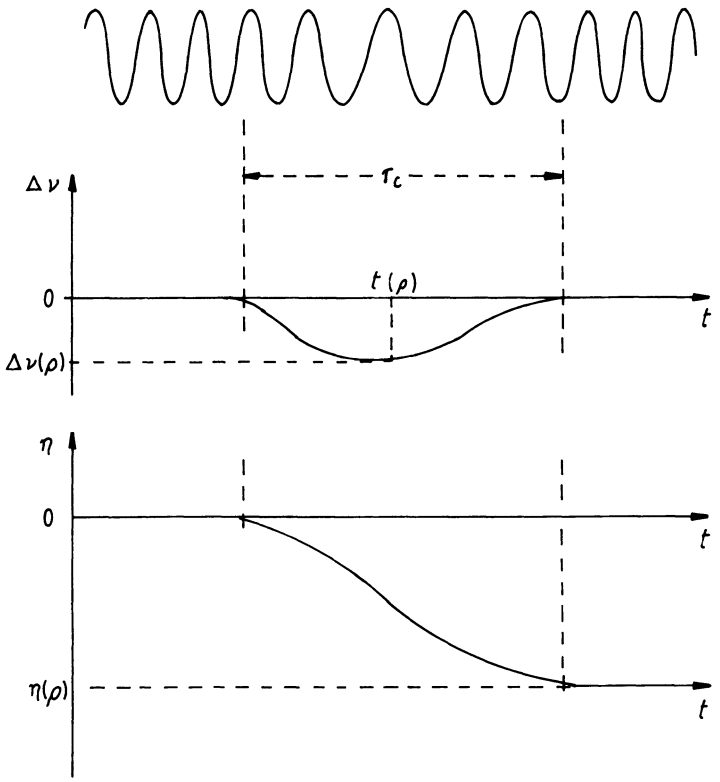

Figure 10.9 Frequency change $\Delta\nu$ and phase change $\eta = \int_0^t 2\pi\Delta\nu \, dt$ as a function of t. The collision duration is τ_c, and the frequency shift has its maximum value when $R = \rho$ at time $t(\rho)$. η is proportional to the area under the $\Delta\nu$ curve and is thus a function of ρ.

The change of limits in the second integral is allowable because $\Delta\nu$ is zero outside the period τ_c.

In the Weisskopf theory a collision producing a phase change of unity is taken to destroy the coherence of the wavetrain and thus to play the part of a quenching collision in the Lorentz theory. The corresponding value of the impact parameter, ρ_W, is known as the Weisskopf radius. No account is taken at this stage of collisions causing smaller phase changes. From the geometry shown in Fig. 10.10,

$$dt = \frac{dx}{v} = \frac{1}{v}\rho \sec^2\theta \, d\theta \quad \text{and} \quad R = \rho \sec\theta$$

Using equations 10.24 and 10.28 for the interaction,

$$\Delta\nu(R) = \frac{C_n(2) - C_n(1)}{hR^n} = \frac{C'_n}{R^n} \quad \text{(say)} \tag{10.29}$$

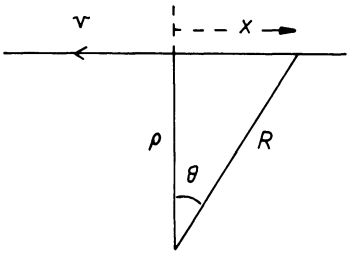

Figure 10.10 Geometry for calculating phase change.

The phase change is then

$$\eta(\rho)=2\pi\int_{-\pi/2}^{\pi/2}\frac{C_n'\cos^{n-2}\theta}{\rho^{n-1}v}\,\mathrm{d}\theta=\frac{2\pi a_n C_n'}{\rho^{n-1}\bar{v}} \tag{10.30}$$

where $\bar{v}$ is the mean relative velocity and a_n is a numerical constant of order unity, depending on the power n. Setting $\eta(\rho)=1$ gives for the Weisskopf radius

$$\rho_W=\left(\frac{2\pi C_n'}{\bar{v}}\right)^{1/(n-1)} \tag{10.31}$$

For the simple Weisskopf model we find t_0 from the number of collisions per second for which $\rho\leqslant\rho_W$,

$$1/t_0=\pi\rho_W^2\,\bar{v}N$$

and use this value of t_0 in equation 10.13 to give for the FWHM of the Lorentzian profile

$$w=\delta\nu_c=\frac{1}{2\pi t_0}=\frac{1}{2}\left(\frac{2\pi C_n'}{\bar{v}}\right)^{2/(n-1)}\bar{v}N$$

This simple model does not work very well for electrons, and it does not give any shift of central frequency. However, it predicts the broadening by neutrals reasonably well, and it also gives the right temperature dependence: for $n=3$, w is independent of v and hence of T, while for $n=6$, $w\propto\bar{v}^{3/5}\propto T^{3/10}$.

The extensions to Weisskopf's theory treat small as well as large phase shifts. The large shifts, resulting from close encounters, are responsible for the width, and the small shifts, from distant encounters, for the central frequency shift. This is readily understandable in a semiqualitative way: a small negative phase shift ($\eta<0$) is equivalent to a slightly smaller *average* frequency over the wavetrain and hence a small red shift (and vice versa for the less usual case of $\eta>0$).

We shall not attempt to present the quantum-mechanical treatment. The basic physics can be understood from Lindholm's semi-classical model, which can be outlined as follows. The problem is to find the frequency power spectrum $I(\omega)$ of an electromagnetic wave $E(t)=E_0\cos\omega_0 t$ which undergoes sudden

phase changes $\eta(\rho)$. One way of doing this is to take the complex Fourier transform of $E(t)$ and find the square of the modulus (as done in section 10.2), but the autocorrelation theorem allows a rather easier route: first take the autocorrelation function of $E(t)$ and then do the Fourier transform. Provided there exists some time interval Δt which is long compared with the duration of a collision but small compared with the time between collisions ($\tau_c < \Delta t < t_0$) the effect of the phase changes is to multiply the unperturbed autocorrelation function $E^2 e^{i\omega_0 t}$ by $e^{\alpha t}$, where α is obtained from the change in the time average of the quantity $(e^{i\eta} - 1)$ over the time Δt. A collision at impact parameter ρ changes the phase by $\eta(\rho)$, and the number of collisions per unit time between ρ and $\rho + d\rho$ is $2\pi\rho\, d\rho\, \bar{v}N$. By the so-called ergodic theorem the time average for a single atom can be replaced by the space average over the ensemble of atoms. Therefore,

$$\Delta(\langle e^{i\eta} - 1\rangle) = \Delta t \int_0^\infty 2\pi\rho\, \bar{v}N(e^{i\eta(\rho)} - 1)\, d\rho$$

and

$$\alpha = 2\pi N\bar{v} \int_0^\infty [(\cos\eta - 1) + i\sin\eta]\rho\, d\rho \tag{10.32}$$

α can conveniently be written in terms of real and imaginary cross-sections

$$\alpha = -N\bar{v}(\sigma_r - i\sigma_i)$$

where

$$\sigma_r = 2\pi \int_0^\infty (1 - \cos\eta)\rho\, d\rho \quad \text{and} \quad \sigma_i = 2\pi \int_0^\infty \sin\eta\, \rho\, d\rho \tag{10.33}$$

It remains to find the spectrum $I(\omega)$ by taking the Fourier transform of $e^{i\omega_0 t} e^{\alpha t}$. Writing this as

$$e^{-N\bar{v}\sigma_r t} e^{i(\omega_0 + N\bar{v}\sigma_i)t}$$

it can be seen that σ_r is responsible for damping, and hence for line broadening, while σ_i gives a shift of central frequency. The Fourier transform is therefore the shifted Lorentzian

$$I(\omega) = \frac{(w/2)^2}{(\omega - \omega_0 - d)^2 + (w/2)^2} \tag{10.34}$$

with FWHM $w = 2N\bar{v}\sigma_r$ and shift $d = N\bar{v}\sigma_i$.

These equations show the different effects of weak and strong collisions. η has the same sign for all perturbers (governed by the sign of C_n'), so for $0 < |\eta| < \pi$, $\sin\eta$ has a non-zero average. Weak collisions therefore contribute to d, but since $\cos\eta \sim 1$ for small η the very weak collisions have little effect on w. For strong collisions, $|\eta| \sim 2\pi$, both the sine and cosine terms average to zero, so $\sigma_i \to 0$ and σ_r approaches the strong collision cross-section $\pi\rho_W^2$.

The condition $\Delta t < t_0$ implies that not more than one collision occurs, in Δt, so that $\eta(\rho)$ can be calculated from equation 10.30 and ρ_W from equation 10.31. The

condition is obviously impossible to satisfy for large enough ρ (weak collisions), so t_0 has to be interpreted as the time interval between strong collisions for which $\rho \leqslant \rho_W$. It is usually assumed that the small phase shifts from weak collisions can be added linearly.

ρ_W has been variously defined by phase changes from 1 to 2π in the course of this discussion, but an exact definition is not important for our purpose, and we can continue to use the original criteria for the impact approximation, equations 10.22 and 10.23. The restriction on number density imposed by equation 10.23, $\rho_W < N^{-1/3}$, is discussed in section 10.5(d) for the different types of perturbation, but it should be noted from equation 10.31 that ρ_W depends on $\bar{v}$ as well as on C_n'. For quadratic Stark broadening, for which $n=4$, $\rho_W^3 \propto 1/\bar{v}$, and the condition $\rho_W^3 < 1/N$ may be satisfied for a given number density of electrons while failing by two orders of magnitude for the same number density of ions.

The other validity condition, equation 10.23, establishes how far out in the wings the impact approximation can be expected to hold. Using the same numbers as in section 10.5(d), one finds a limit on $|\sigma - \sigma_0|$ of 3–10 cm^{-1} ($|\nu - \nu_0| < 100$–300 GHz) for van der Waals and ionic broadening and something under 1 cm^{-1} for resonance broadening. For electrons the impact regime extends over the far wings of the line in virtually all conditions.

Unlike quasi-static broadening, impact broadening actually gives rather little information on interatomic potentials. There are only two parameters to fit, w and d, and the ratio between these is, in principle, determined by the type of interaction potential C_n'/R^n. For $n=6$, $w/d = 2.8$, while for $n=4$, $w/d = 1/\sqrt{3} \simeq 0.6$. The experimental ratios in both cases often turn out to be larger than those predicted by the phase shift model. This may reflect the inadequacy of the multipole expansion, or it may be the result of inelastic collisions, which contribute to the width but not to the shift. These collisions constitute a breakdown of the adiabatic approximation, and a brief discussion is needed.

When a charged particle passes an atom, the latter sees a transient electric field lasting for a time $\tau_c \sim \rho/v$. The principal Fourier components of this disturbance have a frequency

$$\nu \sim 1/\tau_c \sim \bar{v}/\rho$$

If this encompasses the energy gap between the excited state E_2 and an adjacent state E_i of the atom the two states are mixed by the perturbation. The criterion for the breakdown of the adiabatic approximation can therefore be written as

$$\bar{v}/\rho \geqslant |E_2 - E_i|/h$$

It is obviously electrons that are important in this respect. For any given v the inequality can be satisfied by taking ρ very small, but ρ approaches ρ_W the elastic collisions dominate the broadening. Inelastic collisions contribute only if they are significant for $\rho > \rho_W$ – i.e. if

$$\bar{v}/\rho_W \geqslant |E_2 - E_i|/h$$

Substituting for ρ_W from equation 10.31, the condition on $\bar{v}$ becomes

$$(\bar{v})^{n/(n-1)} \geqslant \frac{(C_n')^{1/(n-1)}}{h} |E_2 - E_i|$$

If $|E_2 - E_1|$ is of the order of the fine-structure splitting, one finds that an electron temperature of a few thousand degrees is sufficient to satisfy the inequality.

In the case of neutral particles, non-adiabatic collisions are best envisaged as recoupling of angular momenta in the quasi-molecule. When the molecule breaks up again, the emitting atom is left in a different fine-structure state. The difference in energy is accounted for by a change in the kinetic energy of the perturber. Actually collisions which change the m-values of degenerate systems, but involve no energy transfer, also come under the heading of non-adiabatic collisions. Their effects are manifested by changes in polarization.

The full quantum-mechanical treatment of impact broadening includes matrix elements of all possible states and therefore encompasses non-adiabatic as well as adiabatic collisions. Broadening coefficients can be quite reliably calculated when wave functions are sufficiently well known. Qualitatively, however, the results obtained from the semi-classical theory are entirely compatible with those of the more rigorous treatment.

10.7 Pressure broadening: the quasi-static approximation

The basic assumptions of this approximation are that the perturbers are almost stationary and that the perturbation is nearly constant over the whole time for which the atom is radiating. There are two steps: first to find the effect of a single perturber–emitter pair and then to average over all pairs.

Much effort has been put into the calculation of the $V(R)$s in modern quantum chemistry, exploiting the availability of high computing power for the extensive numerical calculations involved. The simple multipole expansion is inadequate for quantitative studies, largely because the radiating and perturbing atoms cannot really be regarded as separate entities where short-range forces are concerned. At atmospheric pressure there is an appreciable probability of finding the perturber actually within the expectation radius $\langle r \rangle$ of the excited electron orbital. For some lines the far wings (corresponding to small R) show undulations that cannot possibly come from a multipole expansion. Nevertheless, it is well worth running through the quasi-static model with the simple R^{-n} potential in order to see what its main features are.

We start therefore by taking the frequency shift from a single perturber to be given by equation 10.29,

$$\Delta\nu(R) = C_n'/R^n$$

As before, a time average of $\Delta\nu$ for a single atom is considered to be equivalent to

an 'ensemble average' – i.e. an average over all atom–perturber pairs at a single time. Since $\Delta\nu$ always has the same sign (usually negative) the averaging results in an asymmetric spectral line with a shift (usually a red shift) of wavelength.

The simplest averaging model is that first used by Kuhn. A single perturber is assumed to be situated somewhere at random within a sphere of radius R_m. This cut-off radius is defined by the limit of validity of the quasi-static approximation and can be taken as $R_m = \rho_W$ for the reasons given at the end of section 10.4. The line intensity at $\nu_0 + \Delta\nu(R)$ is proportional to the probability of finding a perturber at R:

$$I(\nu)\,\mathrm{d}\nu \propto 4\pi R^2\,\mathrm{d}R$$

Substituting for R from $R^n = C_n'/\Delta\nu$ gives

$$I(\nu) \propto (\Delta\nu)^{-(n+3)/n}$$

For $n=6$ (the van der Waals interaction) this becomes $I(\nu) \propto (\Delta\nu)^{-3/2}$. The assumption of a single perturber is satisfied only if ρ_W is small compared with the average interparticle distance, which is simply the binary interaction condition given in equation 10.23. The extension to multiple perturbers is straightforward for van der Waals interactions, which add in a scalar fashion. For resonance and Stark effects the usual procedure is to start from the 'nearest neighbour' approximation and develop correction terms as a series expansion. It may also be necessary to take account of correlation effects in the case of charged perturbers.

Much of the interest in quasi-static pressure broadening by neutrals attaches to the close connection between the exact form of the interatomic potentials and the details of the line shape. Fig. 10.11 shows typical potential curves similar to those of Fig. 10.8, together with the frequency shift $\Delta\nu$ corresponding to the difference potential. The bonding is sufficiently weak in both states, $|V_{max}| < kT$, that the molecule is temporary and no bound state exists. There is nevertheless a statistically greater chance of finding a perturber near the potential minimum of the initial state, V_2, say, and this can be taken into account by including a Boltzmann factor in the expression for $I(\nu)$:

$$I(\nu)\,\mathrm{d}\nu \propto R^2 \mathrm{e}^{-V_2(R)/(kT)}\,\mathrm{d}R \quad \text{or} \quad I(\nu) \propto \frac{R^2 \mathrm{e}^{-V_2(R)/kT}}{(\partial\nu/\partial R)} \tag{10.36}$$

There is a singularity in this intensity distribution at the value of $\Delta\nu$ for which $\mathrm{d}\nu/\mathrm{d}R = 0$. An infinite intensity at this frequency shift can be avoided by introducing appropriate damping, but it might still be expected that a local maximum would be found. This is indeed the case. Satellites on the broadened wing (usually the red wing) of alkalis perturbed by the inert gases were discovered many years ago and have since been found in other systems. The *position* of the satellite relative to the main peak, $\Delta\nu(R_0)$, is given by the *depth* of the minimum at R_0, and the condition of its appearance is that $\Delta\nu(R_0)$ should exceed the impact-broadened line core. The intensity of the satellite may be

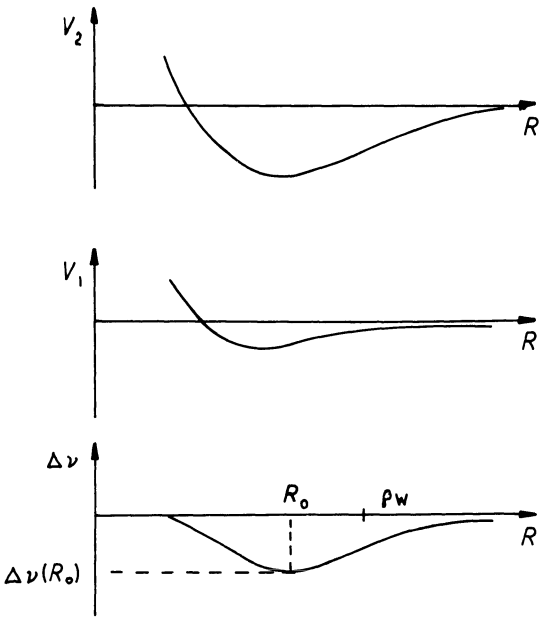

Figure 10.11 Typical potential curves for the upper (V_2) and lower (V_1) states of a transition. The difference potential $\Delta v/h$ has a minimum at R_0, and if $R_0 < \rho_W$, as in the figure, this minimum may give rise to a satellite.

expected to depend on both the depth of the minimum and the temperature by way of the Boltzmann factor in equation 10.36.

The situation for charged particle broadening is a little different because a plasma contains both ionic and electronic perturbers. As has been seen, the impact approximation is usually valid over the whole line for the electrons. As a consequence the criterion for the validity of the quasi-static approximation for the ions can be relaxed, because the smearing of each 'ionic field shift' by ion motion is small compared with that from electron motion. The problem may be tackled by calculating the line shape from the ion fields and then considering each element of the line to be broadened and shifted by the electron impacts. The combined effect is asymmetric broadening for the quadratic Stark effect and symmetric (and much greater) broadening for hydrogen and the H-like ions subject to the linear effect.

It may also be necessary to include in the treatment forbidden transitions induced by the ion fields. As pointed out in section 2.9, external electric fields mix configurations of different parity, so transitions such as s–s and s–d can appear. The ion microfields can have the same effect, and any such 'forbidden' neighbours of the allowed line have to be included in the calculation of the line profile of the latter.

10.8 Experimental aspects of pressure broadening

Experimental work on broadening by neutral atoms has been recently reviewed in a comprehensive article [5] which contains references to earlier reviews and to original papers. The experimental aspects of charged particle broadening can be found in [10, 11], and extensive experimental data are given in [12]. It is not possible to do more than make a few general remarks here.

There are considerable experimental difficulties associated with obtaining a true pressure-broadened line profile. The problem of disentangling the instrument function from the true line shape can be effectively dismissed in situations where laser absorption spectroscopy is possible, but it may still demand a spectrometer of high resolution in the case of emission lines. Having acquired a true line profile, one still has to take into account Doppler broadening and possible self-absorption and there can always be problems with background or shot-to-shot reproducibility.

Most of the experimental work on broadening by neutral atoms has been done on alkali metal lines broadened by inert gases. For experiments carried out in conditions to which the theory can be expected to apply, the line shape, the shift/width ratio, and the temperature and density dependences agree well with the predictions for the most part. Helium tends to be something of a maverick perturber because it has such a small C_6 value that the repulsive term tends to dominate (leading to a blue shift), and classical path assumptions tend to break

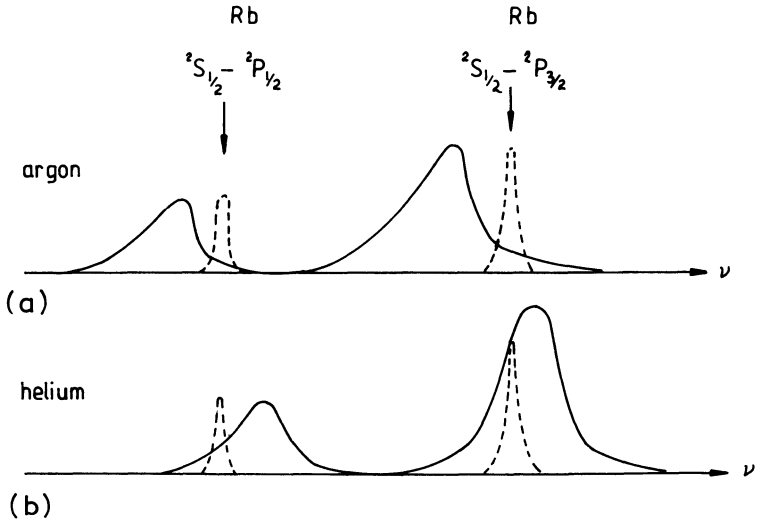

Figure 10.12 Rubidium lines broadened by (a) argon and (b) helium, after [6]. The dotted curves show the doublet at low pressure, and the full curves show broadening by 12 atm. of inert gas.

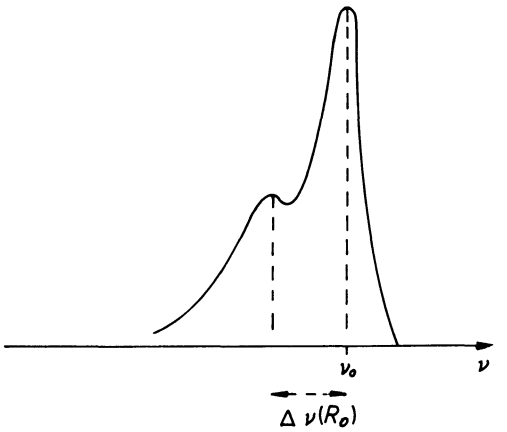

Figure 10.13 Satellite formation in a cesium line broadened by neon, after [5].

down at the very small values of R involved. However, in most experiments the information on interatomic potentials obtained from the shapes of line wings, including any shoulders or satellites, has accorded with what is known of these potentials from other sources and has furnished much additional information. Figure 10.12 shows an example of rubidium lines broadened by (a) argon and (b) helium, and Fig. 10.13 shows a typical satellite formation.

Resonance broadening has been studied in resonance lines themselves and in lines ending on the upper level of a resonance transition. The predictions of zero shift, independence of temperature and linear dependence on density have been confirmed, and the absolute values of the broadening constants agree with theory to within a few percent.

In the case of broadening by charged particles, an immense amount of theoretical and experimental work has been done for the purposes of studying rate processes in plasmas and of developing spectroscopic methods of measuring temperatures and densities. Hydrogen and H-like ions, in particular, are much used for diagnostic purposes because the broadening is very large (linear Stark effect) and well understood. Figure 10.14 shows how the profile of H_β for ion broadening only (Holtsmark profile) is modified when electron impacts and correlation effects are included.

10.9 Convolution of line profiles: the Voigt profile

In the frequent cases when a pressure-broadened line can be represented by a Lorentzian, the combination of pressure and natural broadening is again a Lorentzian, with damping constant $\gamma = \gamma_N + \gamma_C$ where in practice $\gamma_N \ll \gamma_C$. Since

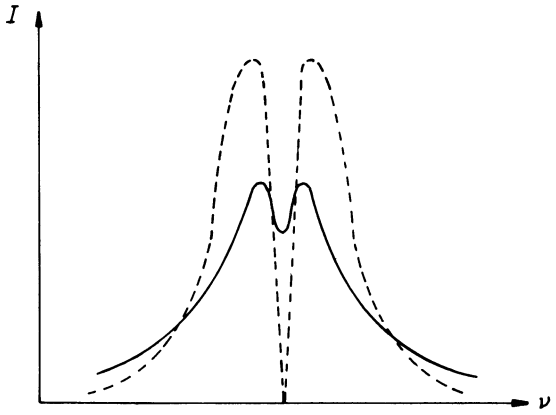

Figure 10.14 Broadening of H_β, after [10]. The broken curve shows the quasi-static ion broadening and the full curve includes the smearing-out effect of electron impact broadening.

Doppler broadening produces a Gaussian profile, very many spectral lines are combinations in some proportion of the Gaussian and Lorentzian line shapes.

If the two types of broadening are independent, the resultant line profile is obtained by 'folding', or convoluting, the two profiles. One can think of each velocity-shifted element as being smeared out into a Lorentzian. The net intensity at any frequency is then the sum of the contributions from the Lorentzians drawn round each Doppler element. The convolution integral is usually written in terms of the reduced frequency x, which is identical with the parameter y defined in equation 10.19, and the damping constant a, which specifies the relative importance of Lorentzian and Gaussian components:

$$I(\mathrm{x}) = \text{const.} \int_{-\infty}^{\infty} \frac{\mathrm{e}^{-y^2}}{(x-y)^2 + a^2} \, \mathrm{d}y \tag{10.37}$$

where

$$x = 2\sqrt{(\ln 2)} \frac{(\nu - \nu_0)}{\delta \nu_{\mathrm{D}}} \quad \text{and} \quad a = \sqrt{(\ln 2)} \frac{\delta \nu_{\mathrm{L}}}{\delta \nu_{\mathrm{D}}}$$

This intensity distribution is known as the Voigt profile and is of considerable practical importance. The integration cannot be done analytically, but it has been computed and tabulated (see, for example, [3,4]) and many codes have been writtten for unfolding it into its component parts. Table 10.1 gives values of $I(x)$ for 5 different damping constants, normalized to keep the area under the curve constant , i.e. $\int \kappa_\nu \, \mathrm{d}\nu = \text{constant}$.

We already know that the Lorentzian may itself be a convolution of two or more Lorentzians, whose FWHM add linearly:

$$\delta\nu_L = \delta\nu_{L1} + \delta\nu_{L2} + \cdots$$

Similarly, the Gaussian may be a convolution of two or more Gaussians, with FWHM added by the rule

$$(\delta\nu_D)^2 = (\delta\nu_{D1})^2 + (\delta\nu_{D2})^2 + \cdots$$

It is often possible to approximate an instrument function by one of these two functions, and so to include it in the convolution.

An important feature of the Voigt function is that the wings are determined almost entirely by the Lorentzian component even when the damping constant a is small. This follows from the exponential decay of the Gaussian: at two full half-widths from the centre a Gaussian has dropped to about 0.2% of its peak intensity, whereas a Lorentzian has dropped to only about 6%. For small a the central part of the Voigt profile and its FWHM are determined essentially by the Gaussian component, and the effect of the Lorentzian is only apparent in the wings. This can be important when one is measuring integrated absorption or total emission intensity in a spectral line, and we shall come back to it when discussing curves of growth in Chapter 11.

Although the Voigt profile is widely used, it is not as universally applicable as is sometimes assumed. Clearly it takes no account of any asymmetry, such as may be found in lines broadened by the quasi-static process. A more insidious trap is that the assumption of independence of the two functions cannot always be justified; impact broadening itself depends on velocity, as seen in section 10.6, and there are other cases of pressure–velocity correlation.

Table 10.1 Normalized Voigt function. The Voigt integral is

$$I(x) = \text{const.} \int_{-\infty}^{\infty} \frac{e^{-y^2}}{(x-y)^2 + a^2}\, dy$$

The table gives values of $I(x)$ as a function of reduced frequency x for different values of damping constant a where

$$x = \frac{2\sqrt{(\ln 2)}}{\delta\nu_D}(\nu - \nu_0), \qquad a = \frac{\sqrt{(\ln 2)}\,\delta\nu_L}{\delta\nu_D}$$

$\delta\nu_D$ is the Doppler FWHM, and the constant is chosen to normalize the integrated intensity to 100 in all cases – i.e.

$$\int_{-\infty}^{\infty} I(x)\, dx = 100$$

x/a	0.01	0.1	0.5	1.0	5.0
0.0	53.36	51.92	35.91	25.14	6.66
0.2	54.15	50.07			
0.4	48.04	44.93	32.73	23.74	6.56
0.6	39.34	37.54			
0.8	29.75	29.27	25.05	20.14	6.45
1.0	20.77	21.34			
1.2	13.39	14.63	16.60	15.65	6.27
1.4	7.98	9.50			
1.6	4.39	5.92	10.08	11.49	6.03
1.8	2.23	3.60			
2.0	1.05	2.19	6.03	8.24	5.75
2.2	0.46	1.38			
2.4	0.19	0.93	3.78	5.96	5.44
		note change of x-step here			
2.8		0.51	2.56	4.41	5.12
3.2		0.35	1.86	3.37	4.78
3.6		0.26	1.42	2.65	4.45
4.0			1.12	2.13	4.13
4.4			0.91	1.75	3.82
4.8			0.76	1.47	3.53
5.2			0.64	1.25	3.27
5.6			0.55	1.07	3.02
6.0			0.47	0.93	2.79
		note change of x-step here			
6.8			0.37	0.72	2.39
7.6			0.29	0.58	2.05
8.4			0.24	0.47	1.78
9.2				0.39	1.55
10.0				0.33	1.36

The table may be extended to larger values of x by using the far wing approximation

$$I(x) \propto \frac{1}{x^2+a^2}$$

For the two extreme cases

(i) $a \to 0$ – pure Gaussian:

$$I(x) = \frac{100}{\sqrt{\pi}} e^{-x^2}$$

Therefore

$$I(0) = \frac{100}{\sqrt{\pi}} = 56.4 \quad \text{and} \quad x_{1/2} = \sqrt{(\ln 2)} = 0.83$$

(ii) $a \to \infty$ – pure Lorentzian:

$$I(x) = \frac{100}{\pi} \frac{a}{x^2+a^2}$$

Therefore

$$I(0) = \frac{100}{\pi a} = \frac{31.8}{a} \quad \text{and} \quad x_{1/2} = a$$

References

INTRODUCTORY TREATMENT

1. Bransden, B.H. and Joachain, C.J. (1983) *Physics of Atoms and Molecules*, Longmans, London.
2. Cowley, C.R. (1970) *The Theory of Stellar Spectra*, Gordon and Breach, New York.
3. Kuhn, H.G. (1969) *Atomic Spectra*, Longmans, London.
4. Mihalas, D. (1978) *Stellar Atmospheres*, Freeman, San Francisco.

BROADENING BY NEUTRAL PARTICLES

5. Allard, N. and Kielkopf, J. (1982) The effect of neutral non-resonant collisions on atomic spectral lines, *Rev. Mod. Phys.*, **54**, 1103.
6. Ch'en, S.Y. and Takeo, M. (1957) Broadening and shift of spectral lines due to the presence of foreign gases, *Rev. Mod. Phys.*, **29**, 20.
7. Hindmarsh, W.R. and Farr, J.M. (1972) Collision broadening of spectral lines by neutral atoms, *Prog. Quantum Electron.*, **2**, 141.
8. Lewis, E.L. (1980) Collisional relaxation of atomic excited states, line broadening and interatomic interactions, *Phys. Rep.*, **58**, 1.
9. Burnett, K. (1985) Collisional redistribution of radiation, *Phys. Rep.*, **118**, 339.

BROADENING BY CHARGED PARTICLES

10. Griem, H.R. (1974) *Spectral Line Broadening by Plasmas*, Academic Press, New York.
11. Griem, H.R. (1975) Stark broadening, in *Advances in Atomic and Molecular Physics* (D.R. Bates and B. Bederson, eds), Vol. 11, p. 331.
12. Konjevic, N., Dimitrijevic, M.S. and Wiese, W.L. (1984) Experimental Stark widths and shifts for spectral lines of neutral atoms, *J. Chem. Phys. Ref. Data*, **13**, 619.
13. As [12] but . . . of positive ions, *ibid.* 649.

ELEVEN

Emission and absorption of line radiation

11.1 Introduction

This chapter is concerned with the total power emitted or absorbed in radiative transitions between discrete energy levels, rather than with the frequency spread discussed in Chapter 10. The power is determined by the relevant transition probability and by the population of the initial state. Both of these quantities have appeared in various forms in earlier chapters. The population of excited states is discussed briefly in section 4.4, and the transition probabilities are introduced in section 2.8 as electric dipole transition moments and picked up again in section 4.5 in the form of the Einstein coefficients. Emission and absorption coefficients j_ν and k_ν appear in Chapter 10, as do radiative lifetimes and a mention of oscillator strength. Chapter 8 also uses these quantities.

The first purpose of this chapter is to derive the relations between all these ways of expressing 'line strengths'. The second purpose is to discuss the effect on the observed radiation of increasing the number of emitting or absorbing atoms in the line of sight – that is, the optical depth. Figure 10.2 shows qualitatively the effect of increasing optical depth on line shape; we shall now tackle this quantitatively, along with the total energy emitted or absorbed and its representation by a curve of growth, a device that is widely used in astrophysics. Finally, we introduce the radiative transfer equations for a homogeneous medium.

11.2 Boltzmann distribution of population and state sum

Boltzmann's formula for the ratio of the populations of two energy levels E_1 and E_2 in thermodynamic equilibrium at temperature T was given as equation 4.17:

$$\frac{N_2}{N_1} = \frac{g_2}{g_1} \mathrm{e}^{-(E_2 - E_1)/kT}$$

where N is usually taken to represent number density and g is the statistical

weight or degeneracy. It is often necessary to relate the population of any given state N_j to the total number density N of atoms of that particular species. If N_0 and g_0 refer to the ground state, we have

$$N = N_0 + N_1 + N_2 + \cdots = \frac{N_0}{g_0}(g_0 + g_1 e^{-E_1/kT} + g_2 e^{-E_2/kT} + \cdots)$$

$$= \frac{N_0}{g_0} \sum_{j=0}^{\infty} g_j e^{-E_j/kT} \quad = \frac{N_0}{g_0} Q(T)$$

where $Q(T)$ is defined by

$$Q(T) = \sum_{j=0}^{\infty} g_j e^{-E_j/kT} \tag{11.1}$$

Therefore $N_0/g_0 = N/Q(T)$, and the population of the jth level is given by

$$N_j = \frac{N}{Q(T)} g_j e^{-E_j/kT} \tag{11.2}$$

$Q(T)$, sometimes designated $U(T)$ or $Z(T)$, is known as the partition function or state sum. The first term is the more usual, and the second the more descriptive.

At moderate temperatures, such that the first excited level has a fairly high energy compared with kT, the partition function converges rapidly, and only a few terms have to be included in the sum. Simple atoms, for example, usually have their resonance lines in the visible or near ultra-violet region, corresponding to $E_1 \sim 25\,000\ \text{cm}^{-1}$. Since $k = 0.7\ \text{cm}^{-1}\ \text{K}^{-1}$, T has to be of the order of 30 000 K for $E_1 \sim kT$, and the populations of the higher levels are very small indeed. However, in complex atoms there may be a number of low-lying states that are appreciably populated even at fairly low temperatures. At very high temperatures the sum to infinity may not converge at all; in such cases the series has to be terminated, a point to which we return in section 13.6.

In molecules one can often ignore the excited electronic states, but the rotational and (usually) the vibrational levels of the ground electronic state have to be included in the state sum. Luckily this does not have to be done level by level, because the sums can be evaluated analytically in terms of the vibrational and rotational constants. From equations 3.12 and 3.9 we have

$$F(J) = BJ(J+1)\,\text{cm}^{-1} \quad \text{and} \quad G(J) = \omega_e(v + \tfrac{1}{2})\,\text{cm}^{-1}$$

The statistical weight of a rotational level with total angular momentum quantum number J is $2J+1$, and it is usually permissible to regard the rotational states as sufficiently close together that the sum may be replaced by an integral. The rotational partition function is then given by

$$Q_{\text{rot}} = \sum_{J=0}^{\infty} (2J+1) e^{-hcBJ(J+1)/kT} = \int_0^{\infty} (2J+1) e^{-hcBJ(J+1)/kT}\, dJ$$

$$= \left[-\frac{kT}{hcB} e^{-hcBJ(J+1)/kT} \right]_0^{\infty} = \frac{kT}{hcB} \tag{11.3}$$

Vibrational levels all have unit statistical weight, giving for the vibrational partition function

$$Q_{\text{vib}} = \sum_{v=0}^{\infty} \mathrm{e}^{-hc\omega_e(v+\frac{1}{2})/kT} \tag{11.4}$$

For the temperature range in which molecules remain undissociated, the population of levels other than $v=0$ is small enough that

$$Q_{\text{vib}} \simeq N_0 = \mathrm{e}^{-\frac{1}{2}hc\omega_e/kT} \tag{11.4a}$$

and it is usually adequate to take

$$N_v \simeq N\mathrm{e}^{-hc\omega_e v/kT}$$

11.3 Einstein coefficients and line strength

The Einstein coefficients are defined in section 4.5. The probability per unit time of an atom in level 2 making a spontaneous transition to level 1 is A_{21}, and the probabilities per unit time of absorption from level 1 to level 2, or of stimulated emission from level 2 to level 1, are $B_{12}\rho(\nu_{12})$ and $B_{21}\rho(\nu_{12})$, respectively, where $\rho(\nu_{12})$ is the spectral density of radiation (J m^{-3} Hz^{-1}) at the frequency of the transition. The relations between the coefficients were found to be (equation 4.19):

$$\left.\begin{aligned} g_1 B_{12} &= g_2 B_{21} \\ A_{21} &= \frac{8\pi h\nu^3}{c^3} B_{21} = \frac{g_1}{g_2}\frac{8\pi h\nu^3}{c^3} B_{12} \end{aligned}\right\} \tag{11.5}$$

These coefficients are intrinsic properties of the atom or molecule and can in principle be calculated if the wave functions are known. To calculate A directly it is necessary to quantize the electromagnetic field as well as the atomic energy states. However, the B-coefficients can be calculated from time-dependent perturbation theory as outlined in section 2.8 (see, for example, [1–4] for a fuller derivation), and A is then found from B by using equation 11.5. The results obtained in this way are identical with those of quantum electrodynamics, so long as the radiation field is weak enough for a perturbation treatment to be valid. (The effects of strong radiation fields are referred to in section 8.7.) In normal conditions we can use equation 2.32 to give for electric dipole transitions

$$B_{12} = \frac{P_{12}}{\rho(\nu)} = \frac{2\pi}{3\hbar^2(4\pi\varepsilon_0)} |\langle 2|er|1\rangle|^2 \tag{11.6}$$

where the modulus term is a shorthand notation for the electric dipole moment averaged over the 'mixed' states (equation 2.33),

$$|\langle 2|er|1\rangle| \equiv |\int u_2^* eru_1 \,\mathrm{d}r|$$

and is made up of the three matrix elements

$$|\langle 2|\boldsymbol{er}|1\rangle|^2 = |\langle 2|ex|1\rangle|^2 + |\langle 2|ey|1\rangle|^2 + |\langle 2|ez|1\rangle|^2 \tag{11.7}$$

No account has been taken of degeneracy so far, and as the subscripts 1 and 2 can be interchanged without altering the value of the integrals equation 11.6 serves for both B_{12} and B_{21}.

The *electric dipole line strength* S of the transition is defined by

$$S_{12} = |\langle 2|\boldsymbol{er}|1\rangle|^2 \tag{11.8}$$

From equations 11.6 and 11.8 we have for a non-degenerate pair of levels

$$B_{12} = \frac{2\pi^2}{3\varepsilon_0 h^2} S_{12}$$

$$A_{21} = \frac{16\pi^3 \nu^3}{3\varepsilon_0 hc^3} S_{12}$$

(Note that $\hbar$ has been replaced by h in these relations.)

The degeneracy must now be considered. If level 1 has degeneracy g_1, the total transition probability $2 \rightarrow 1$ is found by summing over all values of m_1, the quantum number specifying the g_1 different sub-levels. Therefore,

$$A_{21} = \frac{16\pi^3 \nu^3}{3\varepsilon_0 hc^3} \sum_{m_1} |\langle 2|\boldsymbol{er}|1\rangle|^2$$

If level 2 is also degenerate, the decay rate from each of the sub-levels must be the same (otherwise their populations would not stay equal). The transition probability is therefore unaltered by summing it over the g_2-values of m_2 and dividing by g_2 to restore the status quo, giving

$$A_{21} = \frac{1}{g_2} \frac{16\pi^3 \nu^3}{3\varepsilon_0 hc^3} \sum_{m_1} \sum_{m_2} |\langle 2|\boldsymbol{er}|1\rangle|^2$$

The point of this curious manoeuvre is that the symmetry of the individual matrix elements ($\langle 2|ex|1\rangle = \langle 1|ex|2\rangle$, etc.) makes it logical to define the line strength in a symmetric form:

$$S = S_{12} = S_{21} = \sum_{m_1, m_2} \{|\langle 2|ex|1\rangle|^2 + |\langle 2|ey|1\rangle|^2 + |\langle 2|ez|1\rangle|^2\} \tag{11.9}$$

The final expressions for the Einstein coefficients are then

$$\left.\begin{aligned} A_{21} &= \frac{16\pi^3 \nu^3}{3\varepsilon_0 hc^3} \frac{1}{g_2} S \\ B_{21} &= \frac{2\pi^2}{3\varepsilon_0 h^2} \frac{1}{g_2} S \\ B_{12} &= \frac{2\pi^2}{3\varepsilon_0 h^2} \frac{1}{g_1} S \end{aligned}\right\} \tag{11.10}$$

These relations will be found in the collection of Table 11.1, with the corresponding numerical relations in Table 11.2. In using them, or comparing them with expressions elsewhere in the literature, one has to beware of tripping over factors of 2π in no fewer than three places – over and above the $4\pi\varepsilon_0$-factor relating to the CGS system. The questions to ask are: (i) Does the expression use ω or ν? (ii) Does it use $\hbar$ or h? (iii) (the most insidious of all) Is the spectral

Table 11.1 Conversion factors for transition probabilities. Read: $A_{21}=(8\pi h\nu^3/c^3)B_{21}$, etc. S_{ed} is electric dipole line strength

	A_{21}	B_{21}	B_{12}	f	S_{ed}
$A_{21}=$	1	$\dfrac{8\pi h\nu^3}{c^3}$	$\dfrac{g_1}{g_2}\dfrac{8\pi h\nu^3}{c^3}$	$\dfrac{g_1}{g_2}\dfrac{2\pi e^2\nu^2}{\varepsilon_0 mc^3}$	$\dfrac{1}{g_2}\dfrac{16\pi^3\nu^3}{3\varepsilon_0 hc^3}$
$B_{21}=$	$\dfrac{c^3}{8\pi h\nu^3}$	1	$\dfrac{g_1}{g_2}$	$\dfrac{g_1}{g_2}\dfrac{e^2}{4\varepsilon_0 mh\nu}$	$\dfrac{1}{g_2}\dfrac{2\pi^2}{3\varepsilon_0 h^2}$
$B_{12}=$	$\dfrac{g_2}{g_1}\dfrac{c^3}{8\pi h\nu^3}$	$\dfrac{g_2}{g_1}$	1	$\dfrac{e^2}{4\varepsilon_0 mh\nu}$	$\dfrac{1}{g_1}\dfrac{2\pi^2}{3\varepsilon_0 h^2}$
$f=$	$\dfrac{g_2}{g_1}\dfrac{\varepsilon_0 mc^3}{2\pi e^2\nu^2}$	$\dfrac{g_2}{g_1}\dfrac{4\varepsilon_0 mh\nu}{e^2}$	$\dfrac{4\varepsilon_0 mh\nu}{e^2}$	1	$\dfrac{1}{g_1}\dfrac{8\pi^2 m\nu}{3e^2 h}$
$S_{ed}=$	$g_2\dfrac{3\varepsilon_0 hc^3}{16\pi^3\nu^3}$	$g_2\dfrac{3\varepsilon_0 h^2}{2\pi^2}$	$g_1\dfrac{3\varepsilon_0 h^2}{2\pi^2}$	$g_1\dfrac{3e^2 h}{8\pi^2 m\nu}$	1

Notes

1. The Bs are defined in terms of energy density per unit frequency interval.

(a) For energy density per unit wavelength interval

$$A_{21}=\frac{8\pi h\nu^5}{c^4}B_{21}=\frac{8\pi hc}{\lambda^5}B_{21},\quad \text{etc.}$$

(b) For isotropic radiance per unit frequency interval

$$A_{21}=\frac{2h\nu^3}{c^2}B_{21},\ \text{etc.}$$

(c) For isotropic radiance per unit wavelength interval

$$A_{21}=\frac{2h\nu^5}{c^3}B_{21}=\frac{2hc^2}{\lambda^5}B_{21},\quad \text{etc.}$$

2. The integrated absorption coefficient is related to the f-value by

$$\int_{\text{line}} k_\nu\,\mathrm{d}\nu=\frac{e^2}{4\varepsilon_0 mc}N_1 f$$

where N_1 is number density of absorbing atoms.

3. The formulae can be converted to CGS units by replacing ε_0 by $1/4\pi$.

Table 11.2 Numerical conversion factors for transition probabilities. A is in s^{-1}, λ in nm, S in $(C\,m)^2$, and f is dimensionless

	A_{21}	f	S_{ed}
$A_{21}=$	1	$\dfrac{6.670\times10^{13}g_1}{g_2\lambda^2}$	$\dfrac{2.819\times10^{73}}{g_2\lambda^3}$
$f=$	$\dfrac{1.499\times10^{-14}g_2\lambda^2}{g_1}$	1	$\dfrac{4.226\times10^{59}}{g_1\lambda}$
$S_{ed}=$	$3.548\times10^{-74}g_2\lambda^3$	$2.366\times10^{-60}g_1\lambda$	1

Other relations

1. $B_{21}=6.01\times10^{4}\lambda^3A_{21}$

where B is in $s^{-1}\,(J\,m^{-3}\,Hz^{-1})^{-1}$.

2. $\displaystyle\int_{\text{line}} k_\nu\,d\nu=2.65\times10^{-6}N_1f$

where k_ν is in m^{-1}, ν in Hz, and N_1 in m^{-3}.

radiation density hidden in the B definition $\rho(\omega)$ or $\rho(\nu)$ – i.e. is it energy per Hz or energy per unit angular frequency?

It is worth comparing these results with the classical oscillator used in section 10.2. For an electric dipole oscillating in the x-direction with angular frequency ω_0, the dipole moment p_x is

$$p_x=ex=ex_0\cos\omega_0t\equiv p_{0x}\cos\omega_0t$$

The dipole radiates energy at the rate given by equation 10.9,

$$-\frac{\partial U}{\partial t}=\frac{2}{3}\frac{1}{c^3(4\pi\varepsilon_0)}\ddot{p}_x^2$$

Differentiating and taking the cycle average in the same way as before, we get for the average rate of radiation:

$$-\frac{\partial U}{\partial t}=\frac{2}{3}\frac{\omega_0^4}{c^3(4\pi\varepsilon_0)}\frac{1}{2}p_{0x}^2$$

This has to be equal to the energy radiated per second per atom, $A_{21}\hbar\omega_0$. Therefore,

$$A_{21}=\frac{1}{3}\frac{\omega_0^3}{c^3\hbar(4\pi\varepsilon_0)}p_{0x}^2=\frac{4\pi^3\nu^3}{3\varepsilon_0c^3h}p_{0x}^2$$

Comparison with equations 11.9 and 11.10 shows that the two expressions are equivalent if the x-component of the classical oscillator, $p_{0x}=ex_0$, is replaced by

$2|\langle 2|ex|1\rangle|$. The factor 2 has appeared rather accidentally because we have used $\cos\omega_0 t$ instead of $\frac{1}{2}(e^{i\omega_0 t}+e^{-i\omega_0 t})$ in the comparison. Obviously the same relations hold for the y- and z-components, and we end up with the expression

$$p_0 = 2|\langle 2|er|1\rangle| \tag{11.11}$$

for the relation between the classical and quantum-mechanical dipole moments.

The selection rules determining the transitions for which the dipole moment does not vanish identically are discussed in section 2.8. It is mentioned there that transitions forbidden by electric dipole radiation may occur by magnetic dipole or electric quadrupole radiation, corresponding to the next terms in the expansion of the interaction of the electromagnetic field with the atom. The analogy with the classical oscillator, for which the order of magnitude of the electric dipole line strength is given by $S_{ed} \sim (ea_0)^2$, allows us to estimate for the other two types of transition

$$S_{md} \sim \mu_B^2 \quad \text{and} \quad S_{eq} \sim (ea_0^2)^2$$

The relations between these line strengths and the corresponding values of A_{21} are given in Table 11.3. The orders of magnitude are brought out more clearly in Table 11.4, where A_{21} is related to the line strengths expressed in the above atomic units so that each S^a is of order unity. It can be seen that

$$\frac{A_{md}}{A_{ed}} \sim 10^{-5} \quad \text{and} \quad \frac{A_{eq}}{A_{ed}} \sim \frac{1}{100\lambda^2}$$

λ in the second case is in nm, giving a factor 10^{-7}–10^{-8} in the visible region. The magnetic dipole transitions discussed in Chapter 9 that involve *nuclear* spin magnetic moment have line strengths smaller by some six orders of magnitude, as noted in that chapter.

Table 11.3 Magnetic dipole and electric quadrupole radiation. In the numerical relations A is in s^{-1} and λ in nm; S_{md} is in $(A\,m^2)^2$ or $(J\,T^{-1})^2$; S_{eq} is in $(C\,m^2)^2$

Magnetic dipole	$A_{21} = \dfrac{1}{g_2}\dfrac{16\pi^3\mu_0\nu^3}{3hc^3}S_{md} = \dfrac{3.136\times10^{56}}{g_2\lambda^3}S_{md}$
Electric quadrupole	$A_{21} = \dfrac{1}{g_2}\dfrac{8\pi^5\nu^5}{5\varepsilon_0 hc^5}S_{eq} = \dfrac{8.346\times10^{91}}{g_2\lambda^5}S_{eq}$

The relations between A, B and f are as in Tables 11.1 and 11.2.

Table 11.4 Conversion factors for line strengths in atomic units

	Atomic unit	A_{21}	f
Electric dipole	$(ea_0)^2$ 7.187×10^{-59} $(\mathrm{C\,m})^2$	$\dfrac{2.026 \times 10^{15}}{g_2\lambda^3} S^{\mathrm{a}}_{\mathrm{ed}}$	$\dfrac{30.37}{g_1\lambda} S^{\mathrm{a}}_{\mathrm{ed}}$
Magnetic dipole	$(\mu_B)^2$ 8.599×10^{-47} $(\mathrm{J\,T^{-1}})^2$	$\dfrac{2.697 \times 10^{10}}{g_2\lambda^3} S^{\mathrm{a}}_{\mathrm{md}}$	$\dfrac{4.043 \times 10^{-4}}{g_1\lambda} S^{\mathrm{a}}_{\mathrm{md}}$
Electric quadrupole	$(ea_0^2)^2$ 2.021×10^{-79} $(\mathrm{C\,m^2})^2$	$\dfrac{1.680 \times 10^{13}}{g_2\lambda^5} S^{\mathrm{a}}_{\mathrm{eq}}$	$\dfrac{0.252}{g_1\lambda} S^{\mathrm{a}}_{\mathrm{eq}}$

The first column gives the appropriate atomic unit for S. a_0 is the first Bohr radius and μ_B is the Bohr magneton. The second and third columns relate A and f to S expressed in atomic units.

11.4 Absorption coefficient

When a collimated beam of light passes through an absorbing medium, the power absorbed in a distance δx is proportional to both δx and the incident flux. The constant of proportionality is defined in section 10.1 as the absorption coefficient k_ν:

$$-\delta I_\nu = k_\nu I_\nu \delta x \quad \mathrm{W\,m^{-2}\,Hz^{-1}} \tag{11.12}$$

I_ν is the spectral flux density or irradiance, and the units of k_ν are m^{-1}. For a thickness l of absorbing medium, the relation between the incident and emerging power (Fig. 11.1) is found by integrating this equation:

$$I_\nu(l) = I_\nu(0)\,\mathrm{e}^{-\int_0^l k_\nu \mathrm{d}x} \tag{11.13}$$

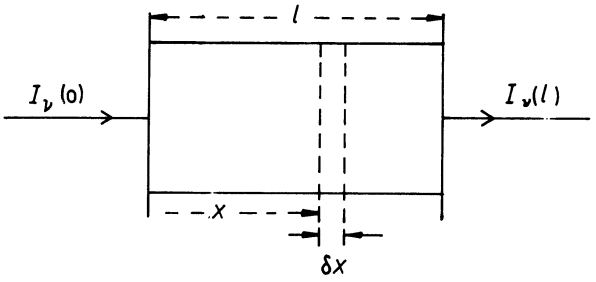

Figure 11.1 Absorption by a column of gas (see text).

$\int_0^l k_\nu \, dx$ is the *optical depth* of the medium. For a homogeneous medium, such that k_ν is independent of x, the optical depth is $k_\nu l$, and we have

$$I_\nu(l) = I_\nu(0)\, e^{-k_\nu l} \tag{11.14}$$

as in equation 10.1. k_ν is the decay constant in the exponential decrease of flux, and $1/k_\nu$ can be thought of as the mean free path of a photon in the medium.

The absorption from an isolated spectral line is spread over a finite frequency range in the ways discussed in Chapter 10. Figure 10.2, repeated here as Fig. 11.2, shows how the apparent shape of the line changes with increasing optical depth. To obtain k_ν as a function of frequency, as in Fig. 11.3, it is necessary to

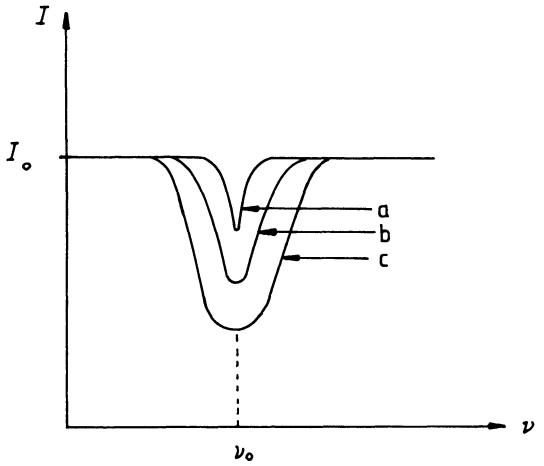

Figure 11.2 Effect of increasing optical depth on line shape. (a), (b) and (c) refer, respectively, to the optically thin, intermediate and optically thick regimes.

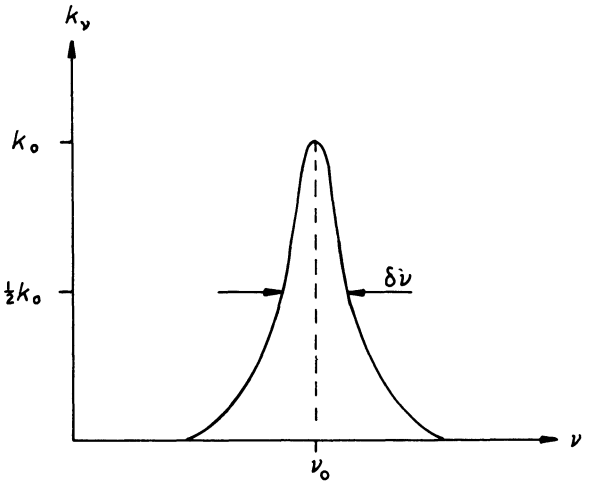

Figure 11.3 Absorption coefficient as a function of frequency.

calculate at each frequency

$$k_\nu = \frac{1}{l} \ln\left(\frac{I_\nu(0)}{I_\nu(l)}\right) \tag{11.15}$$

The simple relation of equation 10.3,

$$I_\nu(l) = I_\nu(0)(1 - k_\nu l)$$

i.e.

$$k_\nu = \frac{1}{l}\left(1 - \frac{I_\nu(l)}{I_\nu(0)}\right) \tag{11.16}$$

is valid *only* if $k_\nu l \ll 1$. When this condition holds the dip in the $I_\nu(l)$ curve is proportional everywhere to $k_\nu l$, and Fig. 11.3 is just curve (a) of Fig. 11.2 upside-down. If terms of order $(k_\nu l)^2$ cannot be neglected, as in (b) and (c), the two curves differ in shape.

While the shape of the k_ν curve depends on the broadening processes involved, the area under it is proportional to the total energy absorbed, which in turn is proportional to the rate of upward transitions $B_{12}\rho(\nu_{12})$. A slab of gas of unit area and thickness δx containing $N_1 \delta x$ ground-state atoms absorbs $N_1 \delta x \rho(\nu_0) B_{12}$ photons per second, where ν_0 is the central frequency of the line. The power absorption is therefore

$$-\Delta I = N_1 \delta x \rho(\nu_0) B_{12} h\nu_0 \quad \text{W m}^{-2}$$

For the collimated beam assumed in this relation and in Fig. 11.1 the flux density is related to the radiation density by $I = \rho c$ (equation 4.9), giving

$$-\Delta I = N_1 \delta x B_{12} I(\nu_0) \frac{h\nu_0}{c}$$

Assuming I_ν in equation 11.12 to have the constant value $I(\nu_0)$ right across the line, the total power absorbed is obtained by integrating k_ν over the line:

$$-\Delta I = I(\nu_0)\delta x \int_{\text{line}} k_\nu \, d\nu$$

Equating these two expressions for ΔI:

$$\int_{\text{line}} k_\nu \, d\nu = N_1 B_{12} \frac{h\nu_0}{c} \quad \text{m}^{-1}\,\text{Hz} \tag{11.17}$$

This expression confirms that the total power absorbed in the line is determined by B_{12} and is independent of the line shape.

Absorption is sometimes expressed in terms of the absorption coefficient per absorbing atom, α_ν. This has the dimensions of area and is often referred to as the (radiative) *atomic cross-section.* Thus

$$\alpha = \frac{k_\nu}{N_1} \text{ m}^2 \quad \text{and} \quad \int_{\text{line}} \alpha_\nu \, d\nu = B_{12} \frac{h\nu_0}{c} \text{ m}^2\,\text{Hz} \tag{11.18}$$

We have so far ignored stimulated emission in these expressions. If there is an appreciable fraction of atoms in the upper state of the transition, photons are put back into the beam, and the net power loss is

$$-\Delta I = \delta x I(\nu_0) \frac{h\nu_0}{c} (B_{12} N_1 - B_{21} N_2)$$

Using equation 11.5 to relate the Bs gives

$$-\Delta I = \delta x I(\nu_0) \frac{h\nu_0}{c} B_{12} N_1 \left(1 - \frac{g_1}{g_2} \frac{N_2}{N_1}\right)$$

The condition $g_1 N_2 > g_2 N_1$ corresponds to population inversion and has been discussed in connection with laser action. For the less drastic degree of upper level population found in thermal sources it is often convenient to work with a *net* absorption coefficient k'_ν defined by

$$k'_\nu = k_\nu \left(1 - \frac{g_1 N_2}{g_2 N_1}\right) \tag{11.19}$$

In terms of k'_ν, by analogy with equation 11.17,

$$\int_{\text{line}} k'_\nu \, d\nu = N_1 B_{12} \frac{h\nu_0}{c} \left(1 - \frac{g_1 N_2}{g_2 N_1}\right) \tag{11.20}$$

11.5 Oscillator strength

The correspondence between classical oscillators and radiating atoms is brought out in the discussion of natural broadening in section 10.2 as well as in section 3 of this chapter. We also use it in section 9.10 in part-explanation of Rayleigh and Raman scattering. The concept of *oscillator strength*, the number of classical oscillators equivalent to one real atom for a particular transition, continues to be a very useful one in the context of absorption, emission, dispersion and scattering of radiation. It is defined by the relation

$$\mathcal{N} = N_i f_{ij} \tag{11.21}$$

where N_i is the number density of atoms (or molecules) in the initial state and $\mathcal{N}$ is the number density of equivalent classical oscillators. f is evidently a dimensionless quantity related to the line strength S and the Einstein coefficients; it is of order unity for a strong allowed transition, and it affords an easy means of comparison between lines of different strength. In this section we shall evaluate the absorption and dispersion predicted by the classical model of an assembly of bound electrons set into damped, forced oscillation by an oscillating electric field and use this, together with equation 11.21, to relate the oscillator strength to the other quantities, B, A and S, governing radiative transitions.

For simplicity we consider oscillations in the x-direction only, produced by a plane-polarized electromagnetic wave travelling in the z-direction. The solution of Maxwell's equations for the electric vector $\boldsymbol{E}$ in a medium of permittivity ε is

$$E(z)=E_0(0)\,\mathrm{e}^{\mathrm{i}\omega(t-z/v)}$$

where the phase velocity v is given by

$$v=\frac{1}{\sqrt{(\varepsilon\mu_0)}}=\frac{c}{\sqrt{(\varepsilon/\varepsilon_0)}}\equiv\frac{c}{n'} \tag{11.22}$$

ε_0 and μ_0 are the permittivity and permeability of vacuum and n' is the complex refractive index, which can be written as

$$n'=\sqrt{(\varepsilon/\varepsilon_0)}=n-\mathrm{i}\kappa \tag{11.23}$$

leading to

$$\frac{z}{v}=\frac{n'}{c}z=\frac{n}{c}z-\mathrm{i}\frac{\kappa}{c}z$$

so

$$E(z)=E_0(0)\,\mathrm{e}^{-(\omega\kappa/c)z}\,\mathrm{e}^{\mathrm{i}\omega(t-nz/c)}$$

Thus κ and n determine the absorption and the phase velocity respectively. κ can be related to the more familiar absorption coefficient k by finding the power attenuation due to the layer z:

$$I(z)\propto|E(z)|^2=E_0^2(0)\,\mathrm{e}^{-2(\omega\kappa/c)z}$$

Comparing this with equation 11.14, we have

$$2\omega\kappa/c=k \tag{11.24}$$

Both n and κ can be determined in terms of the properties of the oscillator by solving the equation of motion for the bound electron driven by the force $-eE(t)$ at some fixed value of z. The equation is

$$\ddot{x}+\gamma\dot{x}+\omega_0^2x=-\frac{e}{m}E_0\,\mathrm{e}^{\mathrm{i}\omega t}$$

where ω_0 is the resonant frequency of the oscillator and γ covers all forms of damping. The equation has the well known solution

$$x=x_0\,\mathrm{e}^{\mathrm{i}\omega t}$$

with

$$x_0=\frac{-(e/m)E_0}{(\omega_0^2-\omega^2)+\mathrm{i}\omega\gamma}$$

The connection to permittivity is made via susceptibility. The instantaneous value of the dipole formed by one electron displaced a distance x is

$$-ex=-ex_0\,\mathrm{e}^{\mathrm{i}\omega t}$$

If there are $\mathcal{N}$ oscillators per unit volume, the susceptibility χ, defined as polarization per unit field, is

$$\chi = -\mathcal{N} e x_0 e^{i\omega t}/E$$
$$= \frac{\mathcal{N}e^2/m}{(\omega_0^2 - \omega^2) + i\omega\gamma}$$

The standard relation between permittivity and susceptibility, $\varepsilon = \varepsilon_0 + \chi$, leads to

$$\frac{\varepsilon}{\varepsilon_0} = 1 + \frac{\mathcal{N}e^2}{\varepsilon_0 m} \frac{1}{(\omega_0^2 - \omega^2) + i\omega\gamma}$$

By equation 11.23 this expression is the square of the complex refractive index n' – i.e.

$$(n')^2 = (n - i\kappa)^2 = 1 + \frac{\mathcal{N}e^2}{\varepsilon_0 m} \frac{1}{(\omega_0^2 - \omega^2) + i\omega\gamma}$$

so

$$(n^2 + \kappa^2) - i2n\kappa = \left(1 + \frac{\mathcal{N}e^2}{\varepsilon_0 m} \frac{\omega_0^2 - \omega^2}{(\omega_0^2 - \omega^2)^2 + \omega^2\gamma^2}\right) - i\left(\frac{\mathcal{N}e^2}{\varepsilon_0 m} \frac{\omega\gamma}{(\omega_0^2 - \omega^2)^2 + \omega^2\gamma^2}\right) \tag{11.26}$$

Equation 11.26 can be put into a simpler form if we are dealing with gases or vapours, for which $n-1$ at STP is of order 10^{-3}–10^{-4}. The approximation $n \simeq 1$ gives κ directly by equating the imaginary parts:

$$\kappa = \frac{\mathcal{N}e^2}{2\varepsilon_0 m} \frac{\omega\gamma}{(\omega_0^2 - \omega^2)^2 + \omega^2\gamma^2} \tag{11.27}$$

For the real parts, we can put $n = 1 + \delta$ and neglect the second-order term δ^2, giving

$$1 + 2\delta + \kappa^2 = 1 + \frac{\mathcal{N}e^2}{\varepsilon_0 m} \frac{\omega_0^2 - \omega^2}{(\omega_0^2 - \omega^2)^2 + \omega^2\gamma^2} \tag{11.28}$$

The second term on the right-hand side is equal to 2δ if κ^2 is also a second-order term. That this is indeed the case, except in the middle of an absorption line, can be seen by comparing this provisional value of δ with κ in the previous equation. Provided $(\omega_0^2 - \omega^2) > \omega\gamma$, $\delta > \kappa$. Therefore

$$\delta = n - 1 = \frac{\mathcal{N}e^2}{2\varepsilon_0 m} \frac{\omega_0^2 - \omega^2}{(\omega_0^2 - \omega^2)^2 + \omega^2\gamma^2} \tag{11.29}$$

with the proviso that this relation may not be valid in the middle of a strong absorption line – where a measurement of refractive index is somewhat impracticable anyway.

The transition from classical oscillators to real atoms is accomplished formally simply by replacing $\mathcal{N}$ by $N_i f_{ij}$ and ω_0 by ω_{ij}, where N_i is the number

density of atoms in the initial state of the transition at ω_{ij} characterized by the oscillator strength f_{ij}. The f-value is thus the effective number of electrons per atom for that particular transition. It is also necessary to sum over all possible transitions from state i. Rewriting the equations in terms of ν instead of ω, we get

$$\left.\begin{aligned} n-1&=\frac{e^2}{8\pi^2\varepsilon_0 m}\sum_j \frac{N_i f_{ij}(\nu_{ij}^2-\nu^2)}{(\nu_{ij}^2-\nu^2)^2+(\nu\gamma/2\pi)^2} \\ \kappa&=\frac{e^2}{8\pi^2\varepsilon_0 m}\sum_j \frac{N_i f_{ij}\nu_{ij}\gamma/2\pi}{(\nu_{ij}^2-\nu^2)^2+(\nu\gamma/2\pi)^2} \end{aligned}\right\} \tag{11.30}$$

Since γ is a measure of linewidth, κ is very small except in the immediate neighbourhood of one of the transition frequencies. Conversely, $n-1$ may collect contributions from transition frequencies far removed from the measurement frequency ν; it is thus a sum over virtual transitions in exactly the same way as the scattering cross-sections discussed in section 9.10.

Equations 11.30 can be further simplified in the spectral region near a particular absorption line, since the term involving this transition, $i\rightarrow k$, say, dominates both sums. Furthermore, we can make the approximations

$$(\nu_{ik}^2-\nu^2)=2\nu_{ik}(\nu_{ik}-\nu) \quad \text{and} \quad \nu\gamma=\nu_{ik}\gamma$$

Near a spectral line, therefore, equations 11.30 take the form

$$\begin{aligned} n_\nu-1&=\frac{e^2}{16\pi^2\varepsilon_0 m}\frac{N_i f_{ik}}{\nu_{ik}}\frac{\nu_{ik}-\nu}{(\nu_{ik}-\nu)^2+(\gamma/4\pi)^2} \\ k_\nu&=\frac{4\pi\nu_{ik}}{c}\,k_\nu=\frac{e^2}{(4\pi\varepsilon_0)mc}N_i f_{ik}\frac{\gamma/4\pi}{(\nu_{ik}-\nu)^2+(\gamma/4\pi)^2} \end{aligned} \tag{11.31}$$

These curves are sketched in Fig. 11.4.

The expression for k_ν is now a Lorentzian, which is to be expected from the starting point of a classical oscillator. f can be related to the Einstein B-coefficient by integrating it to give the total area under the absorption curve in Fig. 11.4:

$$\int_0^\infty k_\nu\,\mathrm{d}\nu=-\frac{e^2N_i f_{ik}}{(4\pi\varepsilon_0)mc}\left[\arctan\frac{\nu_{ik}-\nu}{\gamma/4\pi}\right]_0^\infty=\frac{e^2N_i f_{ik}}{(4\pi\varepsilon_0)mc}\left(\frac{\pi}{2}+\arctan\frac{4\pi\nu_{ik}}{\gamma}\right)$$

Since ν_{ik} is much greater than the linewidth $\gamma/2\pi$ the second term is also $\pi/2$. Therefore,

$$\int_{\text{line}} k_\nu\,\mathrm{d}\nu=\frac{e^2}{4\varepsilon_0 mc}N_i f_{ik} \tag{11.32}$$

All these equations can be expressed in CGS units by replacing $e^2/4\pi\varepsilon_0$ by e^2.

Although equation 11.32 was obtained by integrating a Lorentzian, the right-hand side does not involve any line shape parameter. It was in fact implicitly

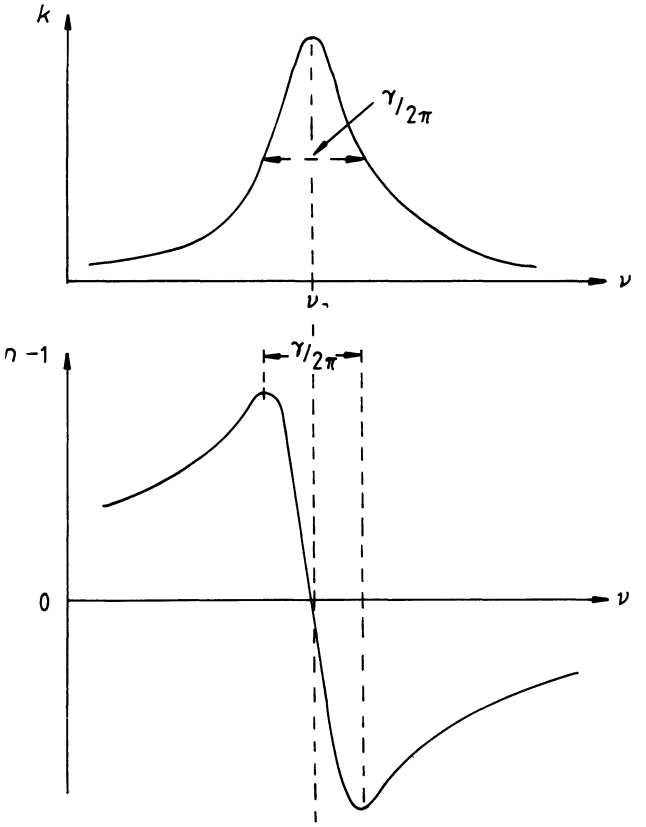

Figure 11.4 Absorption coefficient k and refractive index n as a function of frequency for a classical oscillator with resonance frequency ν_0 and damping constant γ.

assumed in the derivation of equation 11.17 that $\int k_\nu \mathrm{d}\nu$ had no line shape dependence. The relation is therefore valid for *any* line shape. The interlinkage of the intensity parameters can be completed by relating f to the Einstein coefficients through these two equations (11.17 and 11.32):

$$f_{12} = \frac{4\varepsilon_0 mh}{e^2} \nu_{12} B_{12} \tag{11.33}$$

Table 11.1 summarizes the relations between the five quantities A_{21}, B_{21}, B_{12}, f and S, and Table 11.2 gives the corresponding numerical conversion factors. $\int k_\nu \mathrm{d}\nu$ is also included in both tables. Finally, Tables 11.3 and 11.4 give conversion factors for magnetic dipole and electric quadrupole radiation and for atomic units.

f has been taken to represent the absorption oscillator strength f_{12}, which is its normal meaning, but one does occasionally come across the emission

f-value, f_{21}, related to f_{12} by

$$f_{21} = -\frac{g_1}{g_2} f_{12} \tag{11.34}$$

The reason for the minus sign becomes apparent in the next section.

The relationships between f and S for molecular bands and lines are considerably more complex. It is not apparent at first sight what, if anything, is meant by the oscillator strength of a band. Molecular transitions are therefore treated in a special postscript to this chapter.

11.6 The *f*-sum rule

From the interpretation of f as the number of electrons per atom that is optically effective, one would expect some connection between the value of f and the number of valence electrons (one for hydrogen and the alkalis, two for the alkaline earths, etc.). Each electron can take part in several different transitions, so the total oscillator strength should be split in some way between several lines. f therefore represents the fraction of the available electrons participating in a particular transition, and the f-values of reasonably strong lines are of order 1–0.01. The convenient size of f makes it a widely used measure of line strength.

The Thomas–Kuhn–Reiche sum rule states that the sum of *all* transitions from a given state should equal the number of 'optical' or valence electrons, z:

$$\sum_u f_{iu} + \sum_l f_{il} + \int_0^\infty f_{i\varepsilon}\, \mathrm{d}\varepsilon = z \tag{11.35}$$

The subscripts iu and il refer to transitions upwards and downwards, respectively, from the initial state i, while $i\varepsilon$ refers to transitions to the continuum (Fig. 11.5). The downward transitions represent stimulated emission. This is the

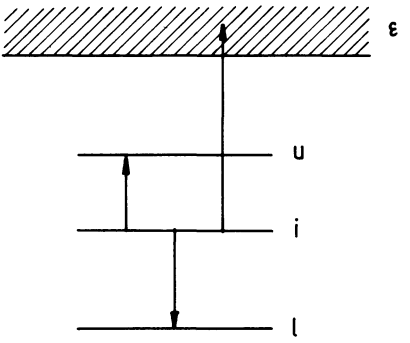

Figure 11.5 Transitions involved in f-sum rule (see text).

reason for the minus sign in equation 11.34: if the downward transitions are expressed in terms of absorption oscillator strengths f_{li} they have to be subtracted from the sum. In terms of absorption oscillator strengths only,

$$\sum_{u>i} f_{iu} - \sum_{l<i} (g_l/g_i) f_{li} + \int_0^\infty f_{i\varepsilon}\, \mathrm{d}\varepsilon = z \tag{11.36}$$

This rule can occasionally be used to put relative oscillator strengths on an absolute scale, but it is less useful than might at first appear. Even if all the absorption f-values to higher bound states are known, the cross-section for continuous absorption has also to be known or estimated; still worse, the virtual transitions to lower levels (which cannot actually take place, because these levels are already full) have to be included in the sum. Partial f-sum rules, relating the f-values for multiplets from a given configuration, are usually more useful in practice.

11.7 Optical depth and equivalent width

The optical depth of a medium is defined in section 11.4 as $\int_0^l k_\nu \,\mathrm{d}x$. It is usually designated by τ_ν (a symbol that has nothing to do with a time constant in this context):

$$\begin{aligned} \tau_\nu &= \int_0^l k_\nu \,\mathrm{d}x \\ &= k_\nu l \quad \text{for a homogeneous medium} \end{aligned} \tag{11.37}$$

A layer of gas is defined as optically thin if $\tau_\nu \ll 1$ and optically thick if $\tau_\nu \gg 1$. Because of the large variations of τ_ν with ν, the layer may be optically thick for one spectral line and thin for another, or thick at the centre of a line but not in the wings. In absorption an optically thin path may be regarded as one in which the absorption does not start to saturate, so that doubling the path length halves the transmitted intensity. In emission it is one in which there is no self-absorption, so that a photon emitted anywhere in the gas is almost certain to get out without being absorbed on the way. The condition $\tau_\nu < 1$ can be interpreted as requiring the mean free path of a photon to exceed the dimensions of the gas layer.

Let us start with the case where the medium is optically thin over the whole profile of an absorption line and examine the effect of increasing the optical depth, by increasing either l or the gas pressure. Figure 11.3 may be regarded as a plot of τ_ν against frequency, and so long as the gas layer is homogeneous, or nearly so, the effect of increasing τ is just to increase the ordinates of this plot without changing the shape. To estimate the optically thin limit, we expand the exponential in equation 11.14,

$$I_\nu(\tau) = I(0)(1 - \tau_\nu + \tfrac{1}{2}\tau_\nu^2 - \cdots) \tag{11.38}$$

Obviously the thickest part of the line is at the centre, and since the optically thin approximation

$$\tau_\nu = 1 - \frac{I_\nu(\tau)}{I_\nu(0)} \tag{11.39}$$

amounts to neglecting the second- and higher-order terms, the criterion for its validity to within, say, 1% is

$$\tfrac{1}{2}\tau_{\nu_0}^2 \leqslant 0.01(1-\tau_{\nu_0}) \Rightarrow \tau_{\nu_0} \leqslant 0.14 \tag{11.40}$$

i.e. 14% peak absorption. In the case of a resonance line ($f \simeq 1$) of a gas or vapour at a pressure of 1 mbar, l has to be of the order of a few μm to satisfy this condition.

Beyond the optically thin limit the transmitted flux no longer decreases linearly with τ, as shown in Fig. 11.2, and the apparent shape of the line changes. I_ν cannot go below the baseline, however large τ becomes, and the absorption saturates first at line centre and then out towards the wings.

The area of the dip in this figure,

$$\int_{\text{line}} [I(0) - I_\nu(\tau)]\, \mathrm{d}\nu$$

is equal to the power lost from the incident beam over the whole line. This quantity, or rather its ratio to the incident power, defines the equivalent width W_ν of the line:

$$W_\nu = \int_{\text{line}} \frac{I(0) - I_\nu(\tau)}{I(0)}\, \mathrm{d}\nu \tag{11.41}$$

The equivalent width is so called because it is the width of the rectangle shown in Fig. 11.6 that has the same area as the actual dip. To a great extent the

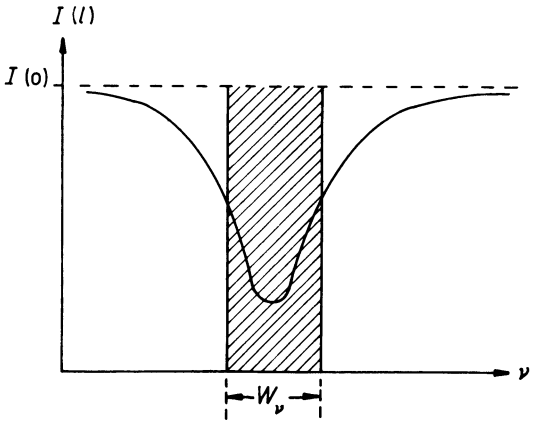

Figure 11.6 Definition of equivalent width. The shaded area is equal to the area between the curve and the $I(0)$ baseline.

equivalent width is independent of instrumental resolution; increasing the slit width of the spectrometer makes the dip broader and shallower but leaves the total area almost unchanged. This property makes it a particularly useful quantity in astrophysics, where low light levels often preclude high resolution. As defined in equation 11.41, equivalent width is evidently measured in frequency units, but it actually masquerades in several slightly different forms: astrophysicists tend to work with wavelengths rather than frequencies and so to measure the equivalent width W_λ in wavelength units, usually ångströms, and both W_ν and W_λ are sometimes multiplied by 2π and referred to as the 'total absorption'.

Equation 11.41 may be rewritten in terms of τ_ν as

$$W_\nu = \int_{\text{line}} (1 - e^{-\tau_\nu}) \, d\nu \tag{11.42}$$

For an optically thin, homogeneous layer this becomes (using equation 11.32)

$$W_\nu = \int_{\text{line}} \tau_\nu \, d\nu = l \int_0^\infty k_\nu \, d\nu = \frac{e^2}{4\varepsilon_0 mc} N_1 f_{12} l \tag{11.43}$$

This important relation is widely used in determining both oscillator strengths and number densities, and it is not infrequently misused in conditions where the optically thin criterion is not fully satisfied. It cannot be too strongly stressed that beyond the optically thin limit the depth of the dip in Fig. 11.2 increases less fast than k_ν, and W_ν therefore increases more slowly than $l \int k_\nu \, d\nu$. The proportionality between W_ν and Nfl is lost. This is not to say that W_ν is no longer a useful quantity; its behaviour through a wide range of optical depth can be described by a 'curve of growth', as discussed in the next section.

11.8 Curve of growth

The curve of growth for a given spectral line describes the behaviour of its equivalent width as the number of absorbing atoms in the line of sight is increased. It usually takes the form of a plot of log W_ν (or log W_λ) against log Nfl. The exact form of the curve depends on the width and shape of the particular line, but the width dependence can be taken care of by replacing W with the dimensionless quantity $W_\nu/\delta\nu_D$ (or $W_\lambda/\delta\lambda_D$) where $\delta\nu_D$ (or $\delta\lambda_D$) is the Doppler FWHM. The result is the set of curves shown in Fig. 11.7, which follow the reproduction in [5–7] of the curves of growth originally computed by van der Held in 1931. These curves assume a homogeneous thickness l of gas, an assumption that may not be justified in a real plasma.

Whatever the line shape, equation 11.43 must hold as long as the absorption path is optically thin. If both sides of the equation are divided by $\delta\nu_D$ it becomes

$$\frac{W_\nu}{\delta\nu_D} = \frac{e^2}{4\varepsilon_0 mc} \frac{Nfl}{\delta\nu_D} = 2.65 \times 10^{-6} \frac{Nfl}{\delta\nu_D} \tag{11.44}$$

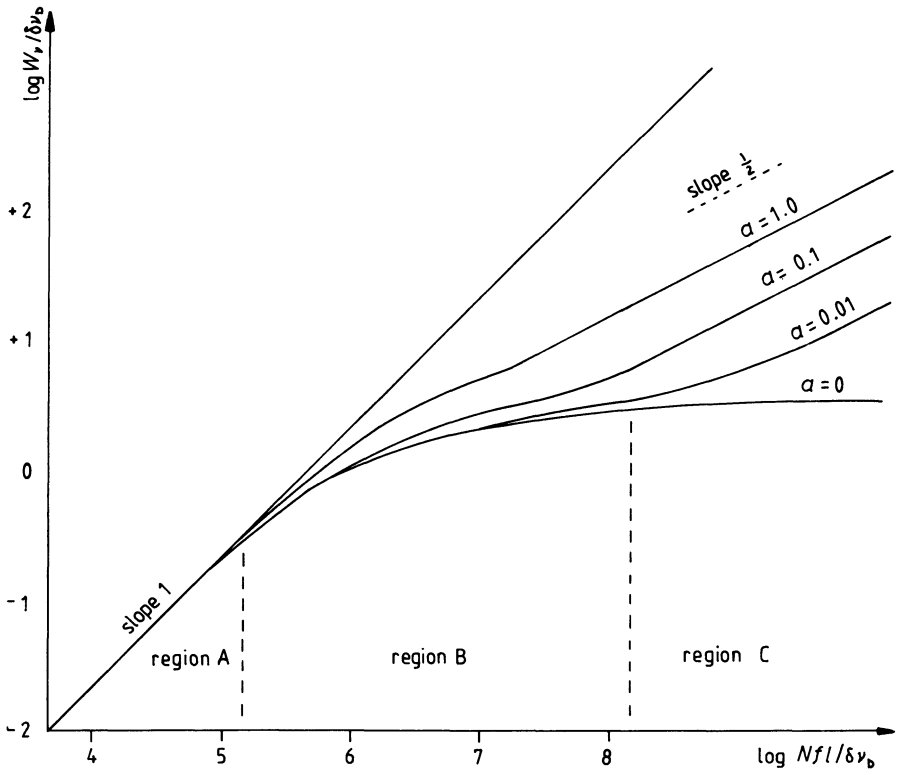

Figure 11.7 Curves of growth for homogeneous column of gas. The argument on the y-axis is the ratio of equivalent width to Doppler FWHM (dimensionless). The numbers on the x-axis are calculated for Nl in m^{-2} and $\delta\nu_D$ in Hz. (If Nl is in cm^{-2}, subtract 4 from each number.) The damping ratio a is defined by $a = \sqrt{(\ln 2)}\,\delta\nu_L/\delta\nu_D$.

where N is in m^{-3}, l is in m and $\delta\nu_D$ is in Hz. Therefore,

$$\log\left(\frac{W_\nu}{\delta\nu_D}\right) = -5.58 + \log\left(\frac{Nfl}{\delta\nu_D}\right) \tag{11.44a}$$

The curve of growth thus starts as a straight line of unit gradient. The absorption in this linear part, marked A on Fig. 11.7, is shown in Fig. 11.8(a).

When the centre of the line is no longer optically thin, W increases less rapidly than Nfl, and the curve starts to flatten out, as in region B. The linear limit and the subsequent behaviour of the curve depend on the spectral line shape. The reason for this is that a Lorentzian profile has much more prominent wings than a Gaussian (section 10.9) and consequently a lower peak height for the same area, as shown in Fig. 11.9. If we define as the linear limit that value of W for which there is a 1% departure from the optically thin approximation, then $W = 0.31\delta\nu_L$ for a pure Lorentzian profile as against $W = 0.15\delta\nu_D$ for a pure

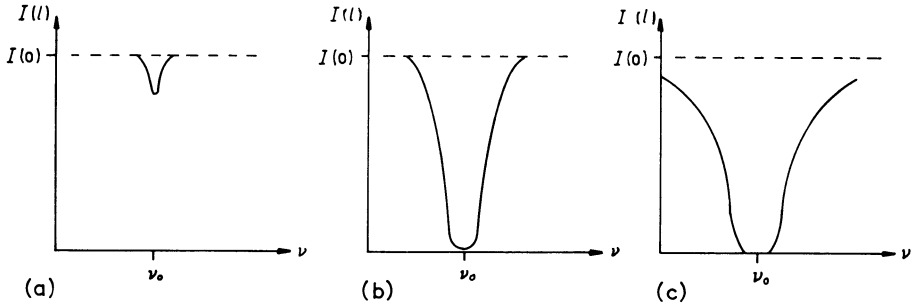

Figure 11.8 Absorption line profiles. The profiles (a), (b) and (c) correspond to the regions A, B and C of the curves of growth in Fig. 11.7.

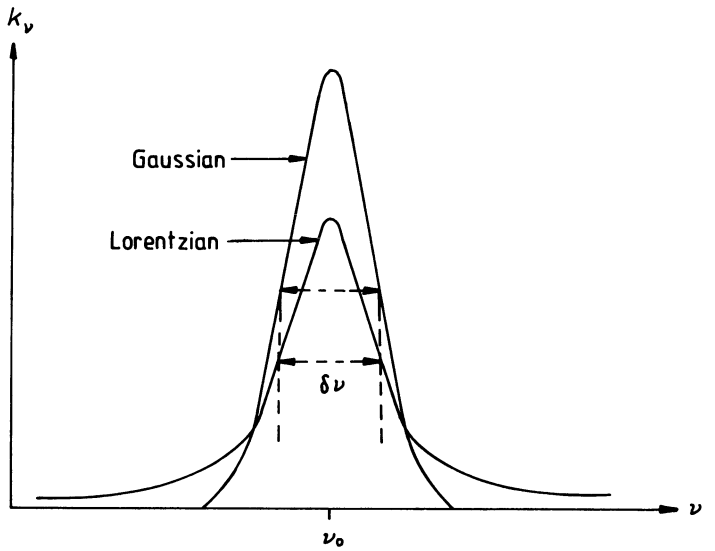

Figure 11.9 Comparison of Doppler and Lorentzian line profiles. The curves have the same area and the same FWHM.

Doppler profile; this is typically 0.01–0.001 cm^{-1}, or a few mÅ. The peak absorption at this linear limit is 20% for a Lorentzian and 15% for a Gaussian. (Note that the 1% error in the *peak* sets in at about 14% peak absorption, as established in equation 11.40, but incipient self-absorption at the peak has a smaller effect on the total area in the case of the Lorentzian.)

The curves of Fig. 11.7 assume a Voigt profile (equation 10.37 and Table 10.1) characterized by the damping ratio a measuring the relative importance of the

Lorentzian and Doppler contributions:

$$a=\sqrt{(\ln 2)}\,\frac{\delta\nu_{\mathrm{L}}}{\delta\nu_{\mathrm{D}}}$$

The characteristics described above are shown quantitatively in Table 10.1: the value of $I(0)$ drops steadily as a is increased, and the contribution of the wings to the total area becomes correspondingly more important.

For the relatively small a values usually met in practice, the central part or core of a Voigt profile is of almost pure Doppler shape, out to perhaps $3\delta\nu_{\mathrm{D}}$ each side of the centre, while the wings are almost pure Lorentzian. The flattish parts of the curve in region B are due to saturation of the Doppler core with very little net increase in W (Fig. 11.8(b)).

If Nfl is increased still further, absorption in the Lorentzian wings starts to become significant and W increases more rapidly, as shown in region C and Fig. 11.8(c). In this region it is again possible to find an analytical expression for W. The absorption in the wings is very closely represented by equation 11.31 with the further approximation $|\nu_{ik}-\nu| \gg \gamma/4\pi$, giving

$$k_\nu=\frac{e^2 N_1 f_{12}}{8\pi\varepsilon_0 mc}\,\frac{\delta\nu_{\mathrm{L}}}{(\nu_{12}-\nu)^2} \tag{11.45}$$

where $\delta\nu_{\mathrm{L}}=\gamma/2\pi$. With this expression for k_ν, $\mathrm{e}^{-k_\nu l}$ can be integrated to give

$$W_\nu=\left(\frac{e^2}{2\varepsilon_0 mc}\,\delta\nu_{\mathrm{L}}\right)^{1/2}(Nfl)^{1/2}$$

or

$$\log(W_\nu/\delta\nu_{\mathrm{D}})=\text{const.}+\tfrac{1}{2}\log a+\tfrac{1}{2}\log(Nfl/\delta\nu_{\mathrm{D}}) \tag{11.46}$$

Consequently the curve of growth finishes up as a straight line of slope $\frac{1}{2}$ whatever the damping ratio, but with lines of different damping ratio displaced relative to one another up the ordinate axis according to the value of $\frac{1}{2}\log a$. For $a=0.1$ the B–C transition occurs when $W \simeq 4\delta\nu_{\mathrm{D}}$. The corresponding Nfl value is about 1000 times greater than the linear limit. For a resonance line of a gas at a pressure of 1 mbar the necessary absorption path is still only a few mm.

Curves of growth have been much used in the determination of f-values (Chapter 12). They also play an important part in astrophysics. In the latter case the assumption of homogeneity is seldom justified, and the curves of Fig. 11.7 do not strictly apply. Special curves have to be calculated for the distribution of temperature, pressure and electron density relevant to the particular model atmosphere used. Measured equivalent widths of lines of a given element can then be compared with these curves. If Boltzmann's formula is valid the abscissa $\log N_1 f_{12} l$ can be written as

$$\log N_1 f_{12}=\text{const.}+\log N+\log g_1 f_{12}-E_1/kT$$

where N is the abundance of the relevant element. The partition function, which is calculable and not very sensitive to temperature, has been incorporated in the

constant. With an appropriate choice of T the points corresponding to measurements on lines of known gf from different initial levels E should all fall on a single curve that can be brought into coincidence with the theoretical curve by sliding it along the abscissa axis. The distance slid gives $\log Nl$ or, since abundances are usually given relative to hydrogen, $\log N/N_{\mathrm{H}}$. Failure to match the curves, or a bad scatter in the points, may indicate that the model atmosphere should be improved or that the laboratory f-values are suspect. Unreliable f-values are, indeed, an important source of uncertainty in work of this kind. More information on the use of curves of growth in astrophysical problems will be found in [5, 6, 8, 9].

11.9 Emission lines; source function and radiative transfer

In the previous section it was tacitly assumed that the layer of gas under study absorbed line radiation without re-emission. This assumption is justified if the absorbing gas is much cooler than the background source, as is the case for light from the interior of a star passing through the stellar atmosphere. In a more general treatment of the transfer of radiation through an absorbing medium both emission and absorption have to be taken into account. A full treatment of radiative transfer may be found in a reference such as [5, 6, 8]; in this section we consider the relatively simple case of line emission from a homogeneous slab of hot gas.

As a starting point, the line emission from an optically thin layer of hot gas is easily found. The number of spontaneous transitions per second in thickness δx and unit cross-section is $A_{21}N_2\delta x$, where N_2 is the number density in the upper state. The energy from each transition is $h\nu_{12}$. The frequency ν_{12} can be assumed to vary little over the line profile and is set equal to the central frequency ν_0. For the moment we neglect stimulated emission. The spontaneous radiation is emitted in all directions, so the power per unit solid angle from unit area – i.e. the radiance – is

$$L(\nu_0)=\frac{1}{4\pi}A_{21}N_2\delta x\, h\nu_0 \tag{11.47}$$

The volume emission coefficient j_ν introduced in section 10.1 is defined properly here as the power emitted per unit solid angle per unit volume per unit frequency interval, or the radiance per unit thickness. Thus, the integral of $j_\nu\delta x$ over the emission line,

$$\delta x\int_{\text{line}} j_\nu \,\mathrm{d}\nu$$

must be equal to $L(\nu_0)$:

$$\int_{\text{line}} j_\nu \,\mathrm{d}\nu=\frac{1}{4\pi}A_{21}N_2 h\nu_0 \tag{11.48}$$

$L(\nu_0)$ is the radiance of the source at ν_0 provided it is optically thin so that virtually all photons emitted escape.

Equation 11.48 is obviously analogous to equation 11.17 connecting $\int k_\nu \, d\nu$ with B_{12}:

$$\int_{\text{line}} k_\nu \, d\nu = N_1 B_{12} \frac{h\nu_0}{c}$$

Taking the two equations together, we have

$$\frac{\int j_\nu \, d\nu}{\int k_\nu \, d\nu} = \frac{c}{4\pi} \frac{N_2}{N_1} \frac{A_{21}}{B_{12}}$$

The right-hand side of this equation is a constant in the sense that it is independent of the line profile. The true line shapes in emission and absorption can be taken as identical for a given transition, so the ratio of the areas on the left-hand side can be replaced by the ratio of the ordinates, j_ν/k_ν, which defines the source function S_ν. Using equation 11.5 for the ratio A/B,

$$S_\nu \equiv \frac{j_\nu}{k_\nu} = \frac{2h\nu_0^3}{c^2} \frac{N_2}{N_1} \frac{g_1}{g_2} \tag{11.49a}$$

If the populations of E_1 and E_2 are in thermal equilibrium so that the ratio N_2/N_1 is given by Boltzmann's relation (equation 4.17), equation 11.49a becomes

$$\frac{j_\nu}{k_\nu} = \frac{2h\nu_0^3}{c^2} e^{-h\nu_0/kT}$$

The right-hand side is Wien's approximation to the blackbody function, equation 4.15:

$$S_\nu = \frac{j_\nu}{k_\nu} = L^W(\nu, T) \tag{11.50a}$$

This result is scarcely surprising, for it is simply an application of Kirchhoff's law relating coefficients of emission and absorption in general to the particular case of line radiation. Wien's relation has appeared instead of Planck's because we have neglected stimulated emission. If for k_ν we substitute the coefficient k'_ν, representing absorption minus stimulated emission – i.e.

$$k'_\nu = k_\nu \left(1 - \frac{N_2}{N_1} \frac{g_1}{g_2}\right)$$

(cf. equation 11.19) – then equation 11.49a is modified to

$$S'_\nu \equiv \frac{j_\nu}{k'_\nu} = \frac{2h\nu_0^3}{c^2} \frac{1}{(g_2/g_1)(N_1/N_2) - 1} \tag{11.49b}$$

which for a Boltzmann population ratio becomes

$$S'_\nu = \frac{j_\nu}{k'_\nu} = L^B(\nu, T) \tag{11.50b}$$

$L^B(\nu, T)$ is now the Planck function of equation 4.12. It must be stressed that this simple relation for the source function is valid only if Boltzmann's relation is valid for the pair of levels concerned, which implies at least partial local thermodynamic equilibrium (see Chapter 13). Alternative expressions for S_ν can be derived for different conditions [5, 6].

We can now investigate the line emission from a column of hot gas as the optical depth is increased beyond the optically thin limit, assuming still that the column is homogeneous. Let the radiance from a length x be $L_\nu(x)$, so that the flux density I_ν in the solid angle $\Delta\Omega$ is given by $I_\nu(x) = L_\nu(x)\,\Delta\Omega$. The addition of a layer of thickness δx (Fig. 11.10) increases the flux density by $j_\nu \delta x \Delta\Omega$, but decreases it by $I_\nu(x) k'_\nu \delta x$. Therefore,

$$\delta I_\nu = (j_\nu \Delta\Omega - I_\nu k'_\nu)\,\delta x$$

or

$$\delta L_\nu = (j_\nu - L_\nu k'_\nu)\,\delta x = (S'_\nu - L_\nu) k'_\nu \delta x$$

For a homogeneous column this expression can be integrated over length l to give

$$[\ln(S'_\nu - L_\nu)]_0^{L(l)} = [-k'_\nu x]_0^l$$

$$\frac{S'_\nu - L_\nu(l)}{S'_\nu} = e^{-k'_\nu l}$$

Therefore

$$L_\nu(l) = S'_\nu(1 - e^{-k'_\nu l}) \tag{11.51}$$

This relation assumes homogeneity, but not necessarily thermodynamic equilibrium. If the latter holds, equation 11.50b can be used to give

$$L_\nu(l) = L^B(\nu, T)(1 - e^{-k'_\nu l}) \tag{11.52}$$

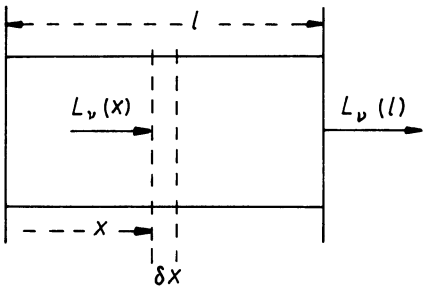

Figure 11.10 Geometry for calculating radiance L from a column of hot gas of length l.

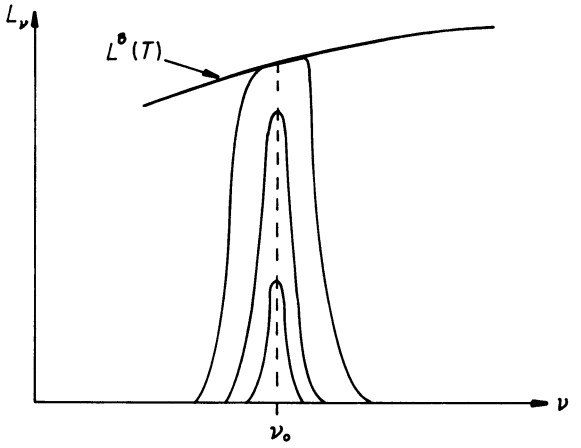

Figure 11.11 Approach to saturation of radiance of gas column as optical depth is increased. $L^B(T)$ is the blackbody curve for temperature T.

As $l \to \infty$, $L_\nu \to L^B(\nu, T)$, the blackbody function for the appropriate ν and T; this is to be expected since a column of large optical depth approaches a blackbody enclosure. Saturation is reached first at the centre of the line and then spreads out towards the wings, behaviour which we have already discussed qualitatively and which is illustrated again in Fig. 11.11. At the opposite extreme, as $k'_\nu l \to 0$,

$$L_\nu(l) \to S'_\nu k'_\nu l = j_\nu l$$

and by integrating j_ν over the line we recover equation 11.47 for the radiation from the complete line.

The saturation of emission of radiation in a gas at constant temperature is not to be confused with the self-absorption and self-reversal that occur if the edges are cooler than the middle. Whereas in the homogeneous case each layer contributes to the emission as well as to the absorption, in the case of a temperature gradient there are fewer excited atoms at the edges, and therefore more absorption than emission. The line peak is first flattened and then reversed, forming a dip in the middle of the line. When emission in the cooler layers is almost negligible, we get back to the case of a pure absorption line. Line shape and intensity can thus be used as a temperature diagnostic, as further discussed in section 13.15.

11.10 *f*-values in diatomic molecules

This section forms a rather lengthy postscript to the discussion of line strengths; it may be omitted without loss of continuity by anyone not concerned with

molecular *f*-values. Molecular transition probabilities are actually discussed briefly in section 3.7, and the reasons for treating the topic at apparently disproportionate length here are, first, that the relations between band and line strengths and oscillator strengths are not immediately obvious and, secondly, that the information is not readily available in standard textbooks. It will be seen that there is ample room for confusion over the statistical weights that enter into the relations between the quantities in Table 11.1. The convention adopted here is that recommended in references [11, 12].

If we start with the nuclei frozen in position, we should have a line strength given as in equation 11.9 by

$$S = \sum_{m_1, m_2} |\int \psi_2^* e r \psi_1 \, \mathrm{d}r|^2$$

where ψ_1 and ψ_2 are the total wave functions. The result of a slow (compared with electron motion) vibrational or rotational motion of the nuclei is not to change significantly the value of this transition moment but to distribute the probability associated with it over a large number of lines; this is equivalent to removing some of the degeneracies in the two sums. To be specific, let us consider absorption; the same arguments can be applied to emission by interchanging the single and double primes throughout. A given rotational level J'' belonging to the vibrational state v'' of the lower electronic state can combine with several different upper vibrational states v', and for each v' there are several possible J'-levels corresponding to the various branches (in the simplest case these are the *P*-, *Q*- and *R*-branches corresponding to $\Delta J = 1, 0, -1$). Ideally we should like to be able to write the strength of one such rotational line $S_{v'v''J'J''}$ as the product of three factors depending, respectively, on the electronic, vibrational and rotational quantum numbers.

The basis of such a factorization is the Born–Oppenheimer approximation introduced in section 3.2: the total wave function is written as the product of electronic, vibrational and rotational functions (equation 3.1),

$$\psi = \psi_e(r, R)\, \psi_v(R)\, \psi_r(\theta, \varphi)$$

The transition moment can then be written (as in equation 3.27) as the product of radial and angular parts

$$\int\int \psi_e'^* \psi_v'^* e r \psi_e'' \psi_v'' r^2 \, \mathrm{d}r \, \mathrm{d}R \equiv R_e^{v'v''}(r, R)$$

and

$$\iint \psi_r'^* \psi_r'' \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\varphi$$

where

$$|\int\int \psi_r'^* \psi_r'' \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\varphi|^2 \equiv S_{J'J''} \tag{11.54}$$

Therefore

$$S_{v'v''J'J''} = |R_e^{v'v''}|^2 \, S_{J'J''} \tag{11.55}$$

The squares of the angular integrals, $S_{J'J''}$, are the Hönl–London factors, which depend on the rotational wave functions only and can be calculated reliably for

virtually all types of coupling in diatomic molecules [10, 14]. The recommended normalization for these factors [12] is

$$\sum_{J'} S_{J'J''} = 2J'' + 1$$

where the sum is taken over all J'-values with which a given J'' can combine. $(2J''+1)$ is the statistical weight of the level J'', corresponding to the $2J''+1$ values of $m_{J''}$. If there is an electronic degeneracy g due to electron spin and/or Λ-doubling (see below), each $J''\rightarrow J'$ transition consists of g lines, and the accepted normalization gives for the sum rule

$$\sum_{J'} S_{J'J''} = g(2J''+1) \tag{11.56}$$

where the sum is taken over all sub-levels of a given J'' and over all upper levels with which they can combine. This electronic degeneracy g is the nub of the confusion, and a little more needs to be said about it.

The electronic degeneracy consists of an electron spin factor $2S+1$ and a Λ-doubling factor 2 for all except Σ-states (section 3.3 and [10]). The degeneracy of a Σ-state is thus $(2S+1)$ and that of a Π-state is $2(2S+1)$. Should the normalization of the Hönl–London factors then be different for $\Sigma\rightarrow\Pi$ transitions and $\Pi\rightarrow\Sigma$ transitions? The answer was 'yes' by the recommendations of 1967 [13], but is now 'no'. The current recommendation [12] is to normalize the electronic part of the transition moment so as to give g in equation 11.56 the same value for upward and downward transitions. Formally, this is expressed as

$$g = (2S+1)(2-\delta_{0,\Lambda'+\Lambda''}) \tag{11.57}$$

but it is probably easier to remember

$$\left.\begin{array}{ll} g=(2S+1) & \text{for } \Sigma\text{–}\Sigma \text{ transitions} \\ g=2(2S+1) & \text{for all other transitions} \end{array}\right\} \tag{11.57a}$$

We now return to the radial part of the transition moment, $R_e^{v'v''}(r, R)$. As the vibrational wave functions do not depend on the co-ordinates r of the electron(s), we can define

$$\int \psi_e'^{*}\, er\, \psi_e'' r^2\, \mathrm{d}r \equiv R_e$$

and write

$$R_e^{v'v''} = \int \psi_v'^{*}\, R_e \psi_v''\, \mathrm{d}R$$

Unfortunately R_e does in general depend on R, so the integral cannot be exactly separated into electronic and vibrational parts. However, R_e is usually assumed to vary sufficiently slowly with R over the range relevant to a particular vibrational transition that it can be taken outside the integral sign to give

$$R_e^{v'v''} = R_e \int \psi_v'^{*} \psi_v''\, \mathrm{d}R \tag{11.58}$$

$\int \psi_v'^{*} \psi_v''\, \mathrm{d}R$ is known as the overlap integral, and its square $q_{v'v''}$ is the Franck–Condon factor. This assumption allows the complete factorization of the

rotational line strength:

$$S_{v'v''J'J''} = |R_e|^2 q_{v'v''} S_{J'J''} \tag{11.59}$$

If the interatomic potential can be expressed in a standard form such as the Morse potential (equation 3.3), the vibrational wave functions and hence the Franck–Condon factors can be computed, giving *relative* band strengths. Vibrational levels have unit statistical weight, and the Franck–Condon factors are normalized to give

$$\sum_{v'} q_{v'v''} = \sum_{v''} q_{v'v''} = 1 \tag{11.60}$$

The calculation of *absolute* line or band strengths demands electronic wave functions, which are much harder to calculate or to estimate reliably.

The recommended definition of the *band strength* is [11]

$$S_{v'v''} = g|R_e^{v'v''}|^2 \simeq g|R_e|^2 q_{v'v''} \tag{11.61}$$

The relation between line and band strengths found by combining this with equation 11.55 is

$$S_{v'v''J'J''} = \frac{1}{g} S_{v'v''} S_{J'J''} \tag{11.62}$$

Experimental measurements lead to transition probabilities in emission and f-values in absorption, so we now need to make sure that the relation between these quantities and the line and band strengths is unambiguous. From Table 11.1, f_{12} and A_{21} are related to S by

$$\left.\begin{aligned} f_{12} &= \frac{1}{g_1}\frac{8\pi^2 m}{3he^2} \nu S \equiv C_1 \frac{\nu}{g_1} S \\ A_{21} &= \frac{1}{g_2}\frac{16\pi^3}{3\varepsilon_0 hc^3} \nu^3 S \equiv C_2 \frac{\nu^3}{g_2} S \end{aligned}\right\} \tag{11.63}$$

where C_1 and C_2 comprise all the constants and g_1, g_2 are the degeneracies of the lower and upper levels, respectively. For a *single line* from a *single sub-level* of the level J'', g_2 is $2J''+1$, so

$$f_{v'v''J'J''} = C_1 \frac{\nu}{2J''+1} S_{v'v''J'J''} = C_1 \nu |R_e|^2 q_{v'v''} \frac{S_{J'J''}}{2J''+1} \tag{11.64}$$

Similarly, the transition probability for a single line from a single sub-level of the level J' is

$$A_{v'v''J'J''} = C_2 \frac{\nu^3}{2J'+1} S_{v'v''J'J''} = C_2 \nu^3 |R_e|^2 q_{v'v''} \frac{S_{J'J''}}{2J'+1} \tag{11.65}$$

A band oscillator strength is a more awkward quantity because of the factor ν in equation 11.63. If the frequency spread over the band is not too large, ν can be set equal to some effective frequency ν_e, and the band oscillator strength $f_{v'v''}$ can be defined by equation 11.63 in terms of the band strength $S_{v'v''}$ as given in

equation 11.61:

$$f_{v'v''} = C_1 \frac{\nu_e}{g''} S_{v'v''} = C_1 \nu_e \frac{g}{g''} |R_e|^2 q_{v'v''} \tag{11.66}$$

where g'' is the electronic degeneracy of the lower state, $2S+1$ for Σ and $2(2S+1)$ for Π, ... Actually the only case for which g and g'' do *not* cancel is a Σ–Π transition. Finally, we can combine this equation with equation 11.64 to relate the line and band oscillator strengths:

$$f_{v'v''J'J''} = \frac{\nu}{\nu_e} \frac{g''}{g} \frac{S_{J'J''}}{2J''+1} f_{v'v''} \simeq \frac{g''}{g} \frac{S_{J'J''}}{2J''+1} f_{v'v''} \tag{11.67}$$

The variation of ν across the band – usually not more than 1–2% – has been ignored in making the last approximation. This equation is an important one in that it allows the entire band absorption to be deduced from the integrated absorption of a single line, on condition only that vibration–rotation interaction is negligible.

In an analogous way, $A_{v'v''}$ is defined as the probability of a transition from v' of the upper electronic state to v'' of the lower one, irrespective of the particular rotational transition. Summing equation 11.65 over all J''-levels accessible from a *single sub-level* of J' gives

$$A_{v'v''} = C_2 \nu^3 |R_e|^2 q_{v'v''} \sum_{J''} \frac{S_{J'J''}}{2J'+1}$$

$$= C_2 \nu^3 |R_e|^2 q_{v'v''} \frac{g}{g'} \tag{11.68}$$

(g' is required in the denominator of the last line because the sum in equation 11.56 is taken over *all* g' sub-levels of J'.) The ν^3 factor means that this expression is independent of J' only insofar as the frequency spread over the band is small. If A is found by taking the band strength from equation 11.61 and applying equation 11.63 with some sort of effective frequency ν_e, the result differs from equation 11.65 by the factor ν_e^3/ν^3 – exactly the same problem encountered with band oscillator strength, but this time raised to the third power.

The frequency spread presents a much worse obstacle to any attempt to define an electronic oscillator strength for a band *system*, which may extend over a hundred nm or so. The quantity f_{el} is still met in the literature, despite efforts to discourage its use. It is defined as the sum of the band oscillator strengths for a given progression:

$$f_{el} = \sum_{v'} f_{v'v''} = \frac{C_1}{g''} \sum_{v'} \nu(v', v'') S_{v'v''}$$

and $\nu(v', v'')$ is usually given some sort of effective mean value $\bar{\nu}$ and brought outside the sum to give

$$f_{el} = \frac{C_1 \bar{\nu}}{g''} \sum_{v'} S_{v'v''} = C_1 \bar{\nu} \frac{g}{g''} |R_e|^2 \sum_{v'} q_{v'v''} = C_1 \bar{\nu} \frac{g}{g''} |R_e|$$

The value of $\bar{\nu}$ depends on the relative weighting of the bands, which is population- and therefore temperature-dependent; f_{el} is really not a very meaningful quantity.

We check finally that the definitions for the line and band f-values are consistent with what we expect from absorption measurements. Suppose there are $N_{v''}$ molecules in the lower vibrational state, distributed among the various rotational levels so that $\sum_{J''} N_{J''} = N_{v''}$. $N_{J''}$ here means the total number in J'' – i.e. there are $N_{J''}/g''$ in each of the sub-levels. The integrated absorption in a line or band is proportional to the total 'Nf-value':

$$W = C_3 N_1 f_{12}$$

where C_3 represents the constants given in equation 11.32 and Table 11.1. If the rotational structure of the band $v'' \rightarrow v'$ is completely unresolved,

$$W = C_3 N_{v''} f_{v'v''}$$

However, if the lines are resolved, we have to sum first over all transitions accessible from a given J'' and then over all values of J'':

$$W = C_3 \sum_{J''} \frac{N_{J''}}{g''} \sum_{J'} f_{v'v''J'J''}$$

$$= C_3 \sum_{J''} \frac{N_{J''}}{g''} \frac{\nu(J'')}{\nu_e} \frac{g''}{g} \sum_{J'} \frac{S_{J'J''}}{2J''+1} f_{v'v''} \quad \text{by equation 11.67}$$

$$= C_3 \sum_{J''} N_{J''} \frac{\nu(J'')}{\nu_e} f_{v'v''} \quad \text{by equation 11.56}$$

$$= C_3 f_{v'v''} \sum_{J''} N_{J''} \frac{\nu(J'')}{\nu_e} \Rightarrow C_3 f_{v'v''} N_{v''} \quad \text{for } \nu(J'') \simeq \nu_e$$

Thus, the two expressions for W agree if the frequency spread over the band is small. If not, the relation $\nu_e N_{v''} = \sum_{J''} \nu(J'') N_{J''}$ logically has to be taken as a definition of ν_e since the physical measurements must give the same values for W. The trouble about defining ν_e in this way is its dependence on the population distribution over $N_{J''}$ and thus on the temperature. Fortunately, in this case, unlike that of the electronic oscillator strength, the ν variation is usually too small to matter. The same arguments apply, of course, to using equation 11.68 to compare the integrated emission from the band with that from the sum over the resolved lines.

References

TRANSITION PROBABILITIES, OSCILLATOR STRENGTHS, ETC.

1. Corney, A. (1977) *Atomic and Laser Spectroscopy*, Oxford University Press.
2. Kuhn. H.G. (1969) *Atomic Spectra*, Longmans, London.

3. Loudon, R. (1973) *The Quantum Theory of Light*, Oxford University Press.
4. Woodgate, G.K. (1970) *Elementary Atomic Structure*, McGraw-Hill, New York.

CURVES OF GROWTH, RADIATIVE TRANSFER, ETC.

5. Cowley, C.R. (1970) *The Theory of Stellar Spectra*, Gordon and Breach, New York.
6. Mihalas, D. (1978) *Stellar Atmospheres*, Freeman, San Francisco.
7. Mitchell, A.C.G. and Zemansky, M.S. (1934) *Resonance Radiation and Excited Atoms*, Cambridge University Press (reprinted 1971).

and (more specifically astrophysical)

8. Aller, L.H. (1963) *The Atmospheres of the Sun and Stars*, Ronald Press, New York.
9. Unsöld, A. (1955) *Physik der Sternatmospharen* (2nd edn), Springer-Verlag, Berlin.

DIATOMIC MOLECULES

10. Herzberg, G. (1950) *Spectra of Diatomic Molecules*, Van Nostrand, New York.
11. Schadee, A. (1978) Unique definitions for the band strength and the electronic vibrational dipole moment of diatomic molecular radiative transitions, *J. Quantum Spect. Radiat. Transfer*, **19**, 451.
12. Whiting, E.E., Schadee, A., Tatum, J.B., Hougen, J.T. and Nicholls, R.W. (1980) Recommended conventions for defining transition moments and intensity factors in diatomic molecular spectra, *J. Molec. Spectrosc.*, **80**, 249.
13. Tatum, J.B. (1967) Interpretation of intensities in diatomic molecular spectra, *Astrophys. J. Suppl.*, **14**, 21.

CALCULATION OF HÖNL–LONDON FACTORS

14. Kovacs, I. (1969) *Rotational Structure in the Spectra of Diatomic Molecules*, Hilger, London.

TWELVE

Experimental determination of transition probabilities and radiative lifetimes

12.1 Introduction

Reliable values for transition probabilities are widely required, particularly in astrophysics and laboratory plasma physics. Most such values are obtained experimentally; calculations tend to be of doubtful reliability unless they have been experimentally vindicated. The experimental methods described in this chapter are, however, of rather wider application than the title indicates, because the same methods can be used to determine parameters of plasmas. For instance, measurements of emission intensities and absorption coefficients yield the product of a population density and an oscillator strength, and if the latter is known the former can be deduced. This aspect is followed up in Chapter 13. The terms 'transition probability' and 'oscillator strength' are used interchangeably, since either can be found from the other by the relations summarized in Tables 11.1–11.4.

Large discrepancies between experimental and calculated values of transition probabilities are not uncommon, and attempts to reduce them have done much towards improving both experimental and theoretical techniques. The existing data for the elements at the beginning of the periodic table (up to calcium, to date) have been collected in the National Bureau of Standards 'Atomic transition probabilities' [1], and these volumes also include brief accounts of, and references to, experimental and theoretical methods. Compilations for certain other elements are listed in [2, 3]. For the rest of the periodic table the information exists in the form of a bibliography, also published by the NBS [4]. Data on oscillator strengths for diatomic molecules are included in the tables of molecular constants in the book by Huber and Herzberg [5].

This chapter covers the most important methods in current use, with brief indications of their relative advantages and limitations. The one thing that all the methods have in common is their inherent uncertainty. Whereas it is

commonplace to measure wavelengths to 1 part in 10^6 with conventional instrumentation and to much higher accuracy by some of the laser techniques, the transition probabilities between the same levels are known to rather better than 1% in a few very favourable cases (the resonance lines of sodium, for example) and to 20% or worse – often considerably worse – in many of the others. A notable feature of the determinations is that results obtained by two different methods, each with a quoted accuracy of say 20%, may differ by a factor of two or more. The NBS compilation classifies an accuracy of 10–20% as 'very good' and excludes only those values for which the error seems likely to exceed 50%. In the case of calculated values it is very difficult even to estimate the probable errors, except for the hydrogen- and helium-like lines, which should be good to better than 1%.

The situation for transition probabilities in molecular spectra is even less satisfactory. Although relative rotational line strengths, and often relative vibrational band strengths, of simple molecules can be calculated with good accuracy, the absolute transition probabilities require a knowledge of the radial part of the electronic wave function, and molecular wave functions are much less accurate than atomic ones. An additional complication arises in that there is often sufficient interaction between the electronic, vibrational and rotational wave functions to make their separation by the Born–Oppenheimer approximation unjustifiable. There is also a considerable dearth of experimental data on molecules, rectified to some extent in recent years by the application of laser spectroscopy to lifetime measurements.

Experimental determinations come under three broad headings: emission, absorption and refractive index, and lifetime. The first two always give the product of a number density and an oscillator strength, and the limitation on accuracy, frequently in the form of a systematic error, is usually the uncertainty in the population. However, it is often possible to obtain very accurate relative oscillator strengths which, ideally, can be put on an absolute scale in conjunction with a lifetime measurement. Similarly, a lifetime measurement does not by itself give an absolute transition probability unless there is only one transition of any strength from the level in question; otherwise measurements (by absorption or emission) of relative oscillator strengths, or 'branching ratios', are required.

The term 'relative oscillator strength' is itself a little ambiguous, for the following reason. The quantity actually measured in emission, absorption or dispersion is $f \int_0^l N_i \, dx$; N_i refers to the upper level of the transition in emission and the lower level in the other two cases, and the integral is the column density, which becomes $N_i l$ for a homogeneous column. A comparison of the oscillator strengths of two lines from the *same* initial level demands no knowledge of N_i and can be made as accurately as the experimental method allows; but relative f-values for lines starting from *different* initial levels require a population ratio, integrated along the line of sight. In practice this can be obtained only from a column in local thermodynamic equilibrium (LTE) with a known temperature

distribution so that Boltzmann's relation (equation 11.2) is valid at every point. The relative column densities are then

$$\frac{\int_0^l N_i' \, dx}{\int_0^l N_i \, dx} = \frac{g_i'}{g_i} \frac{\int e^{-E_i'/kT}}{\int e^{-E_i/kT}} \quad \text{or} \quad \frac{g_i'}{g_i} e^{-(E_i' - E_i)/kT}$$

where the last term holds for a homogeneous column. The greater the energy separation of the initial levels, the more sensitive to T and to departures from LTE does this ratio become. Conditions for LTE are discussed in section 13.8; the principal requirement is a high collision rate, which in turn demands a high particle density, particularly of electrons.

Absolute f-values are still more demanding: the absolute number density of the species, the partition function and the effective column length must all be known. Fortunately, an increasing collection of good relative values has made feasible in many cases the scaling to an absolute value from a single absolute f-value or lifetime measurement. Section 12.5 shows how such a set of relative f-values can be obtained free from assumptions about LTE.

The long-established experimental methods are described in Mitchell and Zemansky's book [12]. The field has been rather well served by review articles over the years, the two latest of which, by Imhof and Read on lifetimes [7] and by Huber and Sandeman on oscillator strengths [6], include the new techniques exploiting lasers, which have been particularly valuable in resolving individual rotational lines in molecular spectra.

12.2 Remarks on the calculation of transition probabilities

A summary of the various methods of calculation is given in the NBS 'Atomic transition probabilities' with further references in the NBS bibliography [1, 4]. The calculation of allowed transition probabilities is discussed in various review articles – e.g. [14, 15, 17] – and forbidden transitions have also been reviewed [16].

Apart from transitions in the hydrogen-like, and to some extent the helium-like spectra, the accuracy of the *ab initio* calculations is typically 10–40%, the lower figure being very difficult to achieve. Semi-empirical calculations, making use of experimental data, may do considerably better than this. In certain cases – when both states of the valence electron may be treated as almost independent of the core – the Coulomb approximation gives satisfactory results. Variational methods based on Hartree–Fock wave functions (the self-consistent field approach with allowance made for electron exchange) are the most generally applicable, but may go badly wrong if there is appreciable configuration interaction. Iso-electronic sequences may be treated by the Z-expansion method, which is essentially a perturbation method using the parameter $1/Z$, Z being the atomic number. This again is affected by configuration interaction.

The calculation of transition probabilities is a much harder task than the

calculation of energy levels because the former are more sensitive to inaccurate wave functions. The results are particularly unreliable if there is a large degree of cancellation in the integrals $\int u_2^* ex u_1 \, d\boldsymbol{r}$ that determine the line strength. Wave functions that are good for small r and bad for large r tend to give good energy values and bad transition probabilities because of the r-weighting in the latter.

The calculation of relative vibrational and rotational transition probabilities in molecular spectra is covered to some extent by the references to Chapter 3, but for electronic transition probabilities a text on quantum chemistry should be consulted.

12.3 Emission measurements

Emission measurements are the most generally applicable, although usually not the most accurate, of the experimental methods. Their basis is given in equation 11.47: the radiance of a spectral line emitted from an optically thin homogeneous layer of thickness l is

$$L(\nu_0) = \frac{1}{4\pi} A_{21} N_2 l h \nu_0$$

where N_2 is the population of the upper state of the transition at ν_0 and A_{21} is the quantity to be determined. The difficulties associated with this apparently simple method are easily summarized.

First, absolute intensity measurements are notoriously difficult to make with high accuracy. The normal practice is to compare the source with a standard source (section 4.3), with due precautions as to solid angles, equivalent light paths, apertures, etc. High temperature sources have necessarily to be compared with standards of very different brightness temperature.

Secondly, the condition for optical thinness may be difficult to satisfy. Most sources have cylindrical symmetry and therefore a radial temperature gradient; viewed side-on, the radiation comes from an inhomogeneous layer, and viewed end-on they tend to be optically thick. The temperature gradient needs to be reliably known if it is to be allowed for properly. Small corrections for self-absorption in a uniform layer can be made, but they tend to be unreliable unless they are indeed small.

Thirdly, there may be continuous emission from the source, and the base line has to be very carefully drawn in order to exclude from the integration both continuum and any scattered light of different wavelength. Moreover, unless the line has a pure Doppler profile the wings extend a very long way out beyond the half-width points. Emission measurements requiring population equilibrium (see below) have to be made at rather high pressures (atmospheric or above) so that a large Lorentzian component to the line profile is almost inevitable. Cutting off the integration as far as five full half-widths from the centre can then introduce an error of $\simeq 7\%$. This makes accurate measurements for close lines extremely difficult and adds to the problems of establishing a true base line.

Fourthly comes the determination of N_2, the problems of which are mentioned in section 12.1. Absolute values of N, the total number density, are very difficult to determine with certainty except for the permanent gases. Vapour pressure measurements tend to be unreliable, and the difficulties of estimating absolute concentrations in a discharge from material introduced in solid or solution form render this method unsuitable for absolute oscillator strengths. Relative measurements, except for branching ratios, still require knowledge of the distribution among the various excited states and therefore the establishment of local thermodynamic equilibrium (LTE). Even with the conditions for LTE fulfilled, considerable errors have been caused by de-mixing effects in arc measurements; the components of the arc plasma tend to separate out, with the more highly excited atoms congregating in the hotter parts of the arc. The necessary corrections for this and other effects are discussed in the more detailed references already given.

Furnaces may be used as emission sources up to about 3000 K; they avoid many uncertainties about equilibrium and temperature, but they are not hot enough for most purposes and may well not be optically thin. Shock tubes seem to be suitable sources, but have been little used in practice. Nearly all measurements have been made with arcs of various kinds. Much of the work done with free-burning arcs is now considered unreliable because of de-mixing, inhomogeneity and deviations from LTE. The 'standard' source at present is the wall-stabilized arc, in which an arc column some 10 cm long and a few mm in diameter runs through holes in a series of water-cooled copper washers. The current may be as high as 100 A, giving axial temperatures up to 13 000–20 000 K and electron densities above 10^{22} m^{-3}. The temperature can be found by plotting log $L_\nu/(g_2 A_{21} \nu)$ against E_2 for lines of known *relative* transition probabilities. If there is true LTE the plot is a straight line of slope h/kT.

As already suggested, emission methods are more suitable for relative than for absolute f-value determinations. Only relative intensity measurements are required, and the need for absolute N-values also vanishes. In the determination of branching ratios, for use in conjunction with lifetime measurements, the equilibrium requirements for the source are also removed since all lines have the same upper level; it is possible to use beam-foil sources and low-current discharges for such measurements. The snag is that a complete set of lines from any one level is required, and this may involve making intensity comparisons between widely different spectral regions.

12.4 Absorption and dispersion measurements

Near an absorption line both the real and the imaginary parts of the refractive index, which determine respectively dispersion and absorption, are proportional to $N_1 f$, where N_1 is the number density in the lower state of the transition of

which f is the oscillator strength (equations 11.31). To obtain absolute f-values from either of these relations it is necessary to know N_1, and as with N_2 for emission measurements this is often the largest source of error, even for lines starting from the ground state. Vapour pressures are notoriously unreliable quantities, and further uncertainties are usually introduced by end-effects at the ends of the column. Absorption from an excited state, other than a very low-lying one, introduces in addition the LTE problems already discussed. Only in the case of relative f-values for lines starting from the same level do all these problems go away. The trick of combining such measurements in emission and absorption to yield a consistent set of relative f-values, independent of population, is described in section 12.6.

This section describes two absorption methods, integrated absorption and equivalent width, and one dispersion method, the hook technique and its variants. These three are the most widely used methods. References [6, 8, 12] should be consulted for variations on these themes – e.g. Faraday rotation and non-linear methods involving lasers – which are applicable to particular types of transition.

12.4.1 INTEGRATED ABSORPTION

This method uses the relation of equation 11.32:

$$\int_{\text{line}} k_\nu \, \mathrm{d}\nu = \frac{e^2}{4\varepsilon_0 mc} N_1 f$$

where k_ν is found as a function of ν from

$$k_\nu = \frac{1}{l} \ln \frac{I_0}{I_\nu}$$

(equation 11.15), and the integrated absorption is the area under the k_ν-curve. The principal difficulty is to resolve the line profile properly. Ideally the instrumental width, or resolution limit, should be considerably less than the true linewidth. If this condition is not fulfilled, the convolution of the instrumental and true line profiles flattens the curve out and gives too low a value for the area. This is basically because incomplete resolution corresponds to a measurement of equivalent width, which is no longer proportional to Nfl once the linear part of the curve of growth is passed (section 11.8). The greater the absorption, the worse is the error; for example, for two-thirds peak absorption ($k_0 l \simeq 2$) the error in $\int k_\nu \, \mathrm{d}\nu$ is 10% for a Doppler-broadened line when the instrumental width is approximately equal to the true width. As the Doppler width of a moderately heavy atom at room temperature is comparable with the resolution limit of a large grating, accurate measurements of integrated absorption require considerable care. One common device is to pressure-broaden the line. Alternatively, adequate resolution can be achieved interferometrically (Fabry–Perot

or Fourier transform spectrometer) or by means of a CW tunable laser scanned through the line profile.

There are two more possible sources of error to be guarded against in both this and the equivalent-width method. One is the presence of scattered light, which reduces the apparent absorption. The other is failure to include the full contribution from the far wings of the line, which is difficult to measure accurately. Both errors result in too low a value for the oscillator strength.

12.4.2 EQUIVALENT WIDTH

The equivalent width W_ν or W_λ defined in equation 11.41 measures the total energy extracted from the incident beam and is independent of the resolution over a wide range. It is the analogue of the emission method and, like this, requires an optically thin column if it is to be used directly in the form of equation 11.43,

$$W_\nu = \frac{e^2}{4\varepsilon_0 mc} N_1 f l$$

corresponding to the linear section of the curve of growth, Fig. 11.7. The range of application can be extended somewhat by use of the curve of growth, as follows. Theoretical curves of growth are plotted for different values of the damping constant a. W is then measured as a function of Nfl, either by increasing the path length or pressure for one line or by using several lines with the same N and known relative f-values (e.g. an atomic multiplet or the rotational lines within a band). If the experimental plot of log W against log Nfl is slid horizontally over the theoretical curves it should eventually fit over one of them. One can then read off both the experimental value of a and the additive constant to log Nfl which gives its absolute value. This fitting procedure is exactly the same as that described in section 11.8 for the determination of astrophysical abundances, except that there lines of known f were used to find N and here known values of N are used to find f. In practice the experimental scatter makes this method difficult to use accurately, particularly if the points fall on the flattish parts of the curve, where a small error in W leads to a large error in Nfl, and it is best regarded as a way of making small corrections to the linear region. An accuracy of 0.5% has been claimed for relative f-values measured in this way with an extremely well-stabilized furnace and the utmost care lavished on all aspects of the experimental conditions and data handling– see [6] for details.

Very strong lines fall in region C, the square root part, of the curve of growth, where

$$W_\nu \propto \sqrt{(\delta\nu_L)}\,\sqrt{(N_1 f l)}$$

If pressure broadening is negligible, so that the Lorentzian contribution to the line is determined entirely by the natural width, the value of $\delta\nu_L$ obtained from

the experimental damping constant,

$$\delta\nu_L = \frac{\delta\nu_D}{\ln 2} a$$

gives an independent estimate of τ_2, the lifetime of the upper state.

12.4.3 HOOK TECHNIQUE

This method of determining the oscillator strength of a line from the anomalous dispersion at its edges was first devised by Roshdestwensky in 1912. It is a particularly convenient and accurate way of using the relationship between n and f given by equation 11.31. If one is concerned with the region 'outside' a line at ν_0–that is,

$$|\nu_0 - \nu| \gg \gamma/4\pi$$

this equation reduces to

$$n - 1 = \frac{e^2}{16\pi^2\varepsilon_0 m\nu_0} \frac{Nf}{\nu_0 - \nu} = \frac{e^2}{16\pi^2\varepsilon_0 mc^2} \frac{Nf\lambda_0^3}{\lambda - \lambda_0} \tag{12.1}$$

The absorbing gas is contained in one arm of a Jamin-type two-beam interferometer (a Mach–Zehnder interferometer is generally used in practice, as shown in Fig. 12.1) illuminated by a continuous source and arranged to give horizontal interference fringes focused on the slit of a stigmatic spectrograph. The path difference between the two beams is very small and is proportional to the distance y up the slit in this configuration, so the fringe maxima are specified by

$$by = m\lambda \tag{12.2}$$

where m is a small integer and b is a geometrical constant; the fringes therefore extend almost horizontally across the spectrum (Fig. 12.2(a)). However, near an

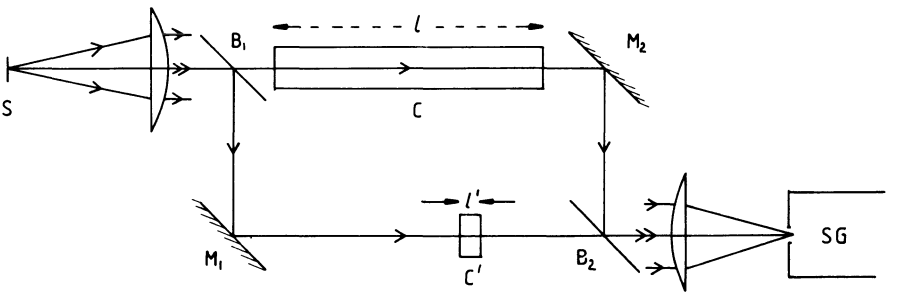

Figure 12.1 Mach–Zehnder interferometer used for 'hook' measurements. B_1, B_2 are beamsplitters and M_1, M_2 are mirrors. C is the gas cell and C′ the 'compensating' plate. The source S and the interference fringes are focused on the slit of the spectrograph SG.

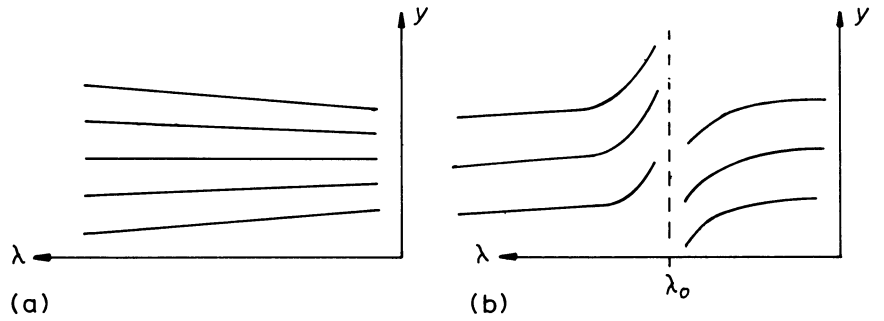

Figure 12.2 Interference fringes (a) far from and (b) close to an absorption line at λ_0. The figures show λ increasing towards the left, so as to conform with the plot of refractive index against λ in Fig. 11.4.

absorption line, the optical path in the top beam changes rapidly with n, and the fringes are curved in the hyperbolic form shown in Fig. 12.2(b). The fringes in fact trace out the $n-1$ curve of Fig. 11.4. In principle the f-value can be found by tracing the path of a given fringe, the locus of which is given by

$$by+(n-1)l=m\lambda$$

In the hook method a 'compensating plate' of thickness l' and refractive index n' is inserted in the lower beam, modifying the fringe equation to

$$by+(n-1)l-(n'-1)l'=m\lambda \tag{12.3}$$

where m is now a large negative integer. The variation of both n and n' with λ can be ignored in any small region clear of absorption lines. The slope of the fringes, found by differentiating equation 12.3 for constant m, is now large:

$$\left(\frac{dy}{d\lambda}\right)_m=\frac{m}{b}$$

Figure 12.3(a) shows these high-order fringes; the direction of the slope depends on b, taken as negative in this figure. Near an absorption line the rate of path increase in the upper beam from the rapidly rising refractive index n must eventually overtake this constant rate, reversing the slope of the fringes. The turning points on either side of the line form the hooks, shown in Fig. 12.3(b). The hook wavelength λ_h is defined by $dy/d\lambda=0$ at constant m. From equation 12.3, with $dn'/d\lambda\simeq 0$, and equation 12.1 we get

$$m=l\frac{dn}{d\lambda}=-l\frac{e^2}{16\pi^2\varepsilon_0 mc^2}\frac{Nf\lambda_0^3}{(\lambda_h-\lambda_0)^2}$$

The two roots of this equation determine the hooks on either side of the line, and

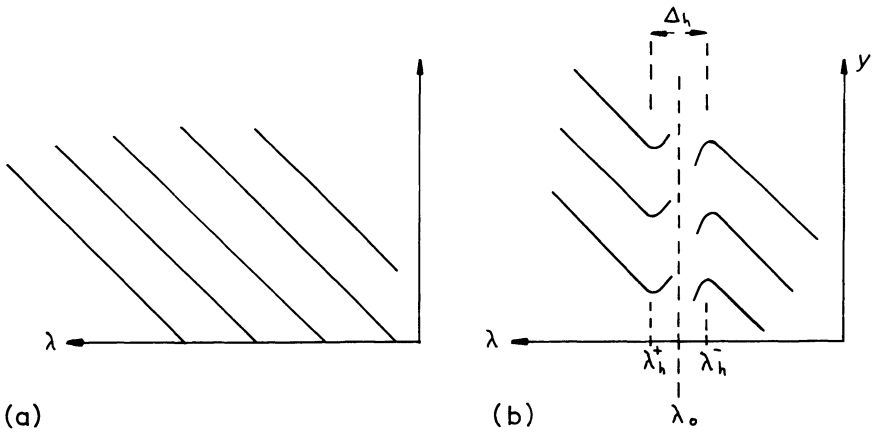

Figure 12.3 High-order interference fringes (a) far from and (b) close to an absorption line at λ_0. Hooks are formed at the two wavelengths λ_h at separation Δ_h.

the distance Δ_h between hooks is given by

$$\Delta_h = 2|\lambda_h - \lambda_0|$$

Putting $-m \equiv K$ gives

$$\Delta_h^2 = \frac{e^2}{4\pi^2 \varepsilon_0 mc^2} \frac{\lambda_0^3 N f l}{K} \tag{12.4}$$

The 'hook constant' K, which is just the negative of the order number, can be found from the undistorted fringes, either by measuring the slope or by counting fringes at constant height across a wavelength interval $\Delta\lambda$:

$$m\Delta\lambda = -\lambda\Delta m \quad \text{or} \quad K = -m = \lambda \frac{\Delta m}{\Delta\lambda}$$

If several close lines are contributing to the refractive index, as in a multiplet, equation 12.1 has to be summed over all relevant lines:

$$n - 1 = \frac{e^2}{16\pi^2 \varepsilon_0 mc^2} \sum_p \frac{N_p f_p \lambda_p^3}{\lambda - \lambda_p}$$

and the hook condition becomes

$$\frac{e^2}{16\pi^2 \varepsilon_0 mc^2} \sum_p \frac{N_p f_p \lambda_p^3}{(\lambda_h - \lambda_p)^2} = \frac{K}{l} \tag{12.5}$$

In general there are $2p$ such equations, corresponding to two hooks for each of p-lines, and the p unknowns $N_p f_p$ can be overdetermined from the solution of $2p$ simultaneous equations. As the hook near any one line depends mainly on the

Nf of that line, with the other lines contributing corrections, the solutions can usually be obtained quite easily by successive approximations. The same method can be applied to the rotational lines in a molecular band, provided that the rotational structure is sufficiently well spaced for hooks to be measured between the lines; if this is not the case, the corresponding roots of equation 12.5 are imaginary. An alternative approach for close rotational structure if the band has a well-defined head is to measure the single hook near the band head; if the *relative* rotational line strengths (the Hönl–London factors of section 11.10) are known, this one measurement determines the band oscillator strength.

The hook method has two great advantages over the absorption methods, and one disadvantage. The disadvantage is the sensitivity: the magnitude of Nfl required for a measurable hook is at least a factor of 10 greater than that required for respectable absorption. The advantages are first that the accuracy does not depend on resolution, optical thinness or any assumptions about line shape and, secondly, that wavelength differences can nearly always be measured more accurately than intensity differences, and the problems of scattered light, absorption from trace impurities, and the far wings of the line are all irrelevant in the hook method.

A number of variants of the basic method have been devised to improve both accuracy and sensitivity. The 'hook vernier' makes use of information from the y-shift of the fringes between the two hook positions either side of the line. The phase shift method measures the increase in fringe separation at a constant value of y in going from the undistorted region towards the line; this method can actually be used with the compensating plate in the same beam as the absorbing gas, in which case the fringe spacing decreases towards the line, but no hooks appear. The phase shift variant is the one most easily adapted to photoelectric rather than photographic measurement, and it can also be used with a narrow-band scanning laser instead of a spectrometer. The additional complexity of these and other variants over the simple hook measurements is justified by their greater sensitivity in the case of weak lines. Further details can be found in [6].

The combination of hook and absorption measurements offers a neat way of determining f-values over a wide dynamic range. Strong lines are 'hooked' in the usual interferometer, and the equivalent widths of weak lines are measured simply by blanking off the lower interferometer beam.

12.5 Lifetime measurements

There are three rather different ways of determining lifetimes of excited states: delay-time measurements, beam measurements and Hanle effect. Accounts of these methods may be found in fairly recent review articles [7, 8] and also in [9–13]. None of them is universally applicable. Some work only for states that can be excited directly from the ground state; others are no good if the lifetime is too short (strong line) or the intensity too low (weak line). Amongst them,

however, they cover a very wide range of transitions, both atomic and molecular, and they have in common the one immensely important advantage that there is no need to know the population of the initial level. A lifetime does not, of course, give directly an oscillator strength, unless there is only one transition from the excited state. The combination of lifetime with relative emission and/or absorption measurements is described in the next section.

12.5.1 DELAY METHODS

If one could excite a large number of atoms to the required level with a very short pulse, of either electrons or radiation, and then measure the intensity of a line starting from this level as a function of time, a decaying exponential would be expected:

$$I(t) \propto A_{21} N(t) = A_{21} N(0) \mathrm{e}^{-t/\tau_2}$$

the time constant of which gives τ_2. The trouble with this rather simple idea is the noise associated with the low photon flux attainable. Fluctuations due to the random arrival of photons follow Poisson statistics, for which the noise is equal to the square root of the signal. Thus, a signal-to-noise ratio of 100 requires at least 10^4 photons per measurement interval. If the lifetime is itself of order 10 ns, a measurement interval can be only a few ns. Taking into account detection efficiency and restrictions on gas density (see below), this method is not usually feasible with conventional methods of excitation, but it certainly can be – and has been – used with tunable laser excitation.

The delayed coincidence method was developed for electron or radiative excitation and is now frequently used with pulsed lasers. The principle of the method is shown schematically in Fig. 12.4. The short exciting pulse also triggers the 'start' of a time-to-pulse-height converter. This is basically a capacitor charged by a steady current, so that the final charge is proportional to the time the current is allowed to flow. The 'stop' signal is provided from the first photon received by the photomultiplier detecting the fluorescence. The pulse height is fed into the appropriate channel of a multichannel pulse height analyser, which displays the result of repetitive pulsing as a histogram of number of counts per channel against channel number, or time delay. The 'channel width' Δt has to be kept small enough to ensure a very small average number of counts per channel, as otherwise the later stop pulses are never recorded. The experiment can be continued until the noise level on the histogram is acceptable, and the time constant is then extracted.

A variant of this method uses modulated instead of pulsed excitation (Fig. 12.5). The emitted radiation is then also modulated, but with a phase shift θ relative to the excitation which depends on the lifetime of the excited state according to $\tan\theta = 2\pi\nu_{\mathrm{m}}\tau$, where ν_{m} is the modulating frequency. The modulation method has been used with both electron and radiative excitation, but the latter is much more common, particularly since lasers have become available.

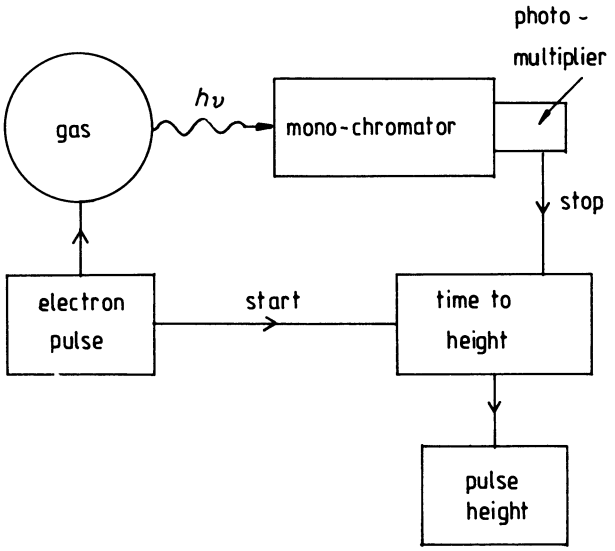

Figure 12.4 Schematic arrangement for lifetime measurement by delayed coincidence. See text for explanation.

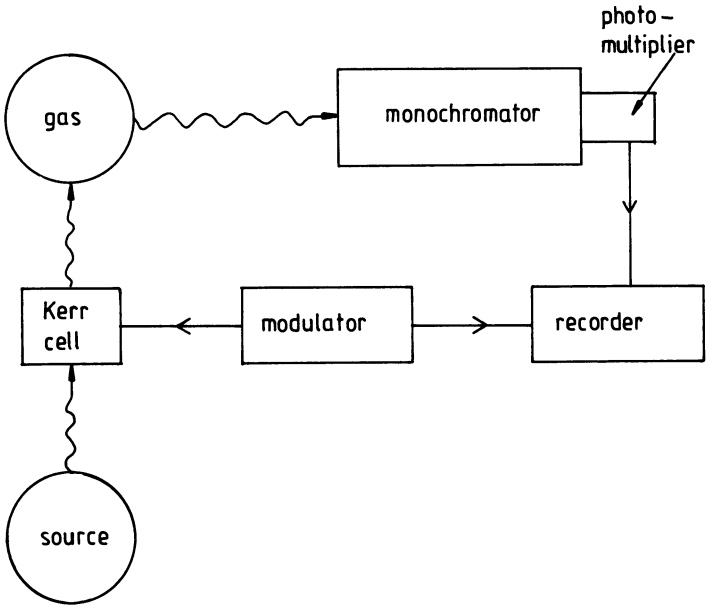

Figure 12.5 Schematic arrangement for lifetime measurement by phase shift. See text for explanation.

The modulation can be done either electro-optically, with Kerr cells, or acousto-optically, with ultrasonic waves.

The delay methods can be used for lifetimes from a few microseconds to a few nanoseconds. Apart from difficulties associated with scattered light and sometimes insufficient spectral resolution, they have three principal potential sources of systematic error. The first, cascading, is relevant only to electron or broadband radiative excitation. If higher levels are excited and subsequently decay to the level under investigation, the lifetime of the latter is apparently prolonged by the repopulation. This difficulty can evidently be avoided by selective population of one level only, a great bonus provided by tunable lasers. The second problem affects all levels emitting to the ground or a metastable state and is known as imprisonment of resonance radiation. If the gas is not optically thin, some of the photons are absorbed and re-emitted one or more times before they eventually get out, and the effect is, again, to prolong the apparent lifetime of the state. Finally, longer-lived states may be collisionally depopulated, or 'quenched', which of course shortens the apparent lifetime. Both of these last problems can be dealt with in principle by going to sufficiently low pressure, but one may then run into difficulties with low light intensities. Systematic errors of this type have often been underestimated in the experimental data.

12.5.2 BEAM MEASUREMENTS

The beam foil light source is described in section 4.8.6. A beam of ions of various degrees of ionization emerges from the thin foil in different states of excitation. The excited states decay as the ions travel downstream from the foil, and the rate of decrease of the intensity of any particular line as a function of distance from the foil gives directly the lifetime of the relevant excited state. Historically this method is the successor to the experiments of Wien in the 1920s on canal rays, in which the lifetime was measured from the decay of emitted radiation as the excited ions in a discharge tube travelled beyond the cathode. The beam foil method has intrinsically much greater accuracy because of the much higher ion velocity. In practice the spectrometer is kept fixed and the intensity is measured as the foil is moved upstream. It is necessary to monitor the constancy of the beam while this is going on, either by measuring the total charge collected at the end or by using a second photomultiplier.

The principal difficulties of the method are cascading from higher excited states, low light intensity and large Doppler broadening. The first of these is covered in the previous section. The second, rather obviously, is a consequence of the small population of a particular level of a particular stage of ionization in a beam that is in any case of low density ($\sim 10^5$ ions cm^{-3}). Photon counting methods are usually necessary. The reason for the Doppler problem is illustrated in Fig. 12.6. The velocity of the beam is of order 5×10^6 m s^{-1}, and because of the finite acceptance angle of the monochromator the direction of observation is not strictly perpendicular to the beam direction. For light

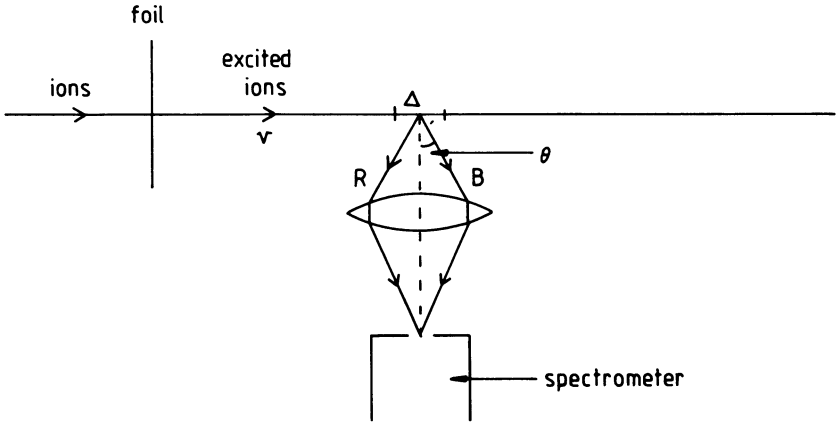

Figure 12.6 Schematic arrangement for beam-foil spectroscopy. The finite acceptance angle θ results in a red shift for light travelling along ray R and a blue shift for light travelling along ray B, and hence in Doppler broadening.

travelling along ray B the ions have a component of velocity along B of $v \sin \theta$, resulting in a blue shift, whereas for ray R the velocity is $v \sin \theta$ away from the monochromator, giving a red shift. As all rays in the cone RB contribute to the image, the result is a broadened line. The low intensity makes a wide aperture essential, and the Doppler broadening may be as high as $10^{-3}\lambda$. Obviously, for isolated lines this does not matter, but the beam foil spectrum tends to be such a rich one that spectral overlaps often present problems.

The density in the beam is so low that there are no difficulties with imprisonment of radiation or collisional quenching. The method is best suited for lifetimes of about 10^{-8}s, during which the ions travel a few cm, but it has actually been used over the whole range 10^{-12}–10^{-7} s. The short lifetimes present problems of spatial resolution because the decay distance is itself so short, whereas the longer lifetimes yield very low intensities.

A powerful alternative to foil excitation has been furnished by tunable lasers. Very accurate lifetime measurements have been made by crossing a laser beam tuned to a particular excitation frequency with a fast-ion beam. Both cascading and spectral overlap problems are eliminated in this technique, and it is even possible to resolve fine-structure components, or rotational structure in molecular ions, and to populate higher excited levels by stepwise excitation.

Much of the importance of the beam methods is that they are applicable to the higher stages of ionization, which are usually beyond the reach of the other methods. Data obtained by following up iso-electronic sequences are important both theoretically and practically and have been used to extrapolate to yet higher degrees of ionization.

12.5.3 HANLE EFFECT

This method is again a resurrection from the 1920s, when the effect, investigated by Hanle, was known as magnetic depolarization of resonance radiation. It is now also known as zero-field level-crossing, being a special case of the level-crossing experiments described in section 9.9 for the measurement of hyperfine structure. For a very brief qualitative description of the effect it is simplest to use a semi-classical model, but when there is hyperfine structure or close fine structure in the excited level a proper quantum-mechanical treatment is required. A typical experimental layout is shown in Fig. 12.7. Light from the source of resonance radiation travelling in the y-direction is plane-polarized in the x-direction before entering the resonance vessel. The emitted radiation is due to electrons oscillating along the x-axis and has the same polarization. A photomultiplier on the x-axis therefore records no signal. If now a magnetic field is applied in the z-direction, the resonance radiation is partly depolarized, and the intensity rises at strong fields to half that in the absence of the polarizer. The depolarization can be ascribed to the precession of the polarization direction about the z-axis with angular frequency ω_B given in terms of the Landé g-factor by

$$\omega_B = g_J \mu_B B/\hbar$$

If ω_B is sufficiently rapid, the components of the oscillation along the x- and y-axes are equalized, and the y-oscillations radiate along the x-axis. The lifetime enters into the story because at small ω_B some of the atoms decay before the precessional cycle is completed, so that the depolarization is only partial. The signal thus depends on the relative magnitudes of the precessional and decay

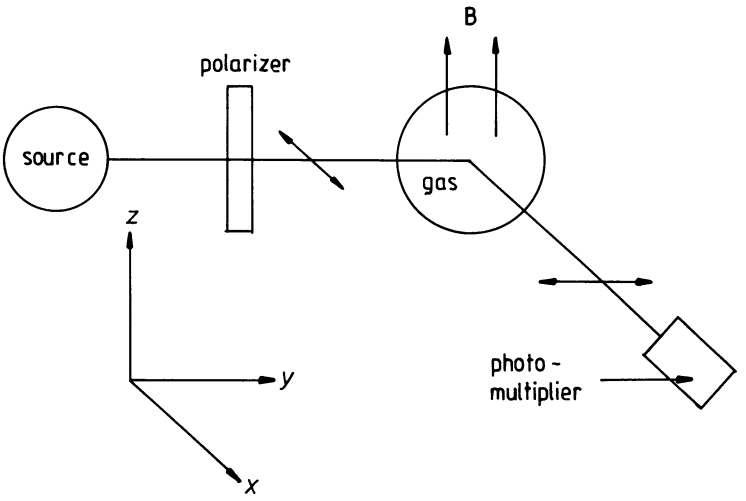

Figure 12.7 Schematic arrangement for Hanle method. See text for explanation.

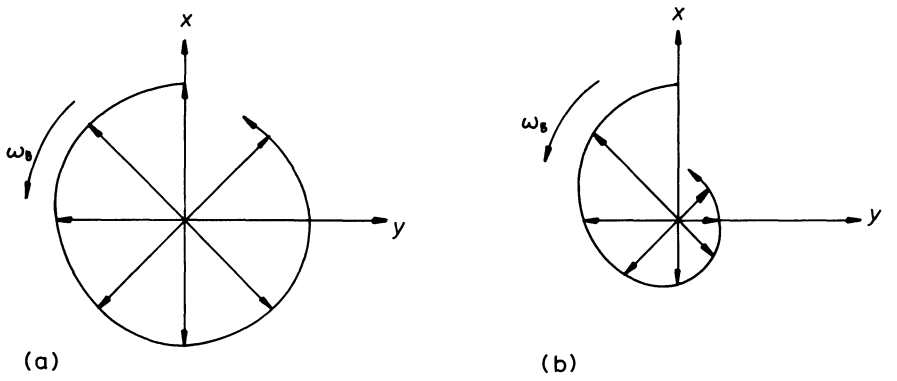

Figure 12.8 Classical interpretation of Hanle effect. (a) represents small damping ($\omega_B > 1/\tau$) and (b) represents large damping ($\omega_B < 1/\tau$). The length of each spoke is proportional to the intensity of radiation for that particular direction of polarization.

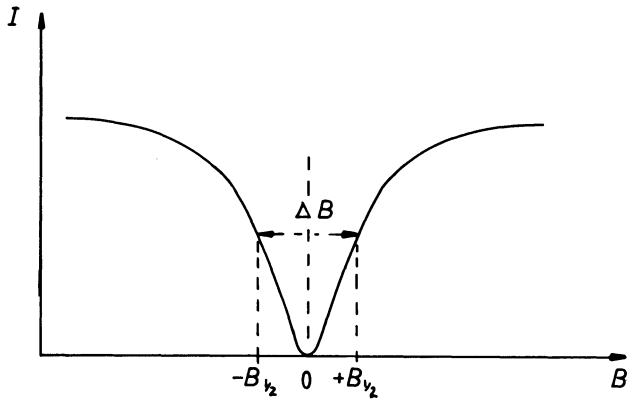

Figure 12.9 Signal as a function of magnetic field in the Hanle effect. The curve is an inverted Lorentzian, and $g_J\tau$ can be found from its FWHM (see text).

times, as shown schematically in Fig. 12.8: in terms of the damping constant $\gamma = 1/\tau$, (a) represents $\omega_B \gg \gamma$ and (b) shows $\omega_B \ll \gamma$. The semi-classical treatment gives for the signal (see, for example [11])

$$I = \text{const.}\left(1 - \frac{\gamma^2}{\gamma^2 + 4\omega_B^2}\right)$$

This expression has the inverted Lorentzian form shown in Fig. 12.9. The signal has half its maximum value when

$$\frac{\gamma^2}{\gamma^2 + 4\omega_B^2} = \frac{1}{2} \quad \Rightarrow \quad \omega_B = \frac{\gamma}{2} = \frac{1}{2\tau}$$

therefore

$$g_J\tau = \hbar/2\mu_B B_{1/2}$$

τ can therefore be determined from the half-width of the signal-versus-field curve if g_J is known – for example, from the closely allied level-crossing or double resonance experiments.

The quantum mechanical interpretation was outlined in section 9.9. In zero field all the magnetic sub-levels of the upper state are degenerate and are populated coherently. The emitted radiation from them interferes destructively in the x-direction. To lift the degeneracy, it is necessary to separate the sub-levels by an amount ΔE_B greater than their natural width $\hbar/\tau$, leading to the condition

$$g_J\mu_B B > \hbar/\tau$$

which is fulfilled beyond the half-width point. Further degeneracies, with consequent changes of polarization, occur whenever the field is such that the energies of two hyperfine components overlap, as already seen in level-crossing spectroscopy. These other degeneracies may have to be taken into account at the fields of order 10^{-3}–10^{-2} T required for the 'zero-field' measurements.

This technique has no cascading or collisional quenching problems, and the 'lines' are free of Doppler and power broadening. Radiation trapping is still a potential source of error, in much the same way as the coherence narrowing described in section 9.8: absorbing atoms in the same magnetic field are in a sense coherent with the emitting atoms, and the polarization information is passed on to them by the photon. However, with due care the method can be made to give accuracies of a few percent for lifetimes in the micro- to nano-second range, and with selective laser excitation it has been applied to many molecular as well as atomic transitions.

12.6 Combinations of methods

Given that excited-state lifetimes can often be measured with considerable accuracy independently of the number density, there is a strong incentive for using them to put relative measurements made by emission or absorption methods on an absolute scale. A lifetime does not in itself give a transition probability except in the few cases where a resonance line is the only transition of any strength from the level in question. If there is more than one allowed transition from the level k, as shown in Fig. 12.10, we have

$$\tau_k = 1/\sum_i A_{ki}$$

and we need relative intensity measurements for all transitions from k to obtain an absolute A-value for any one of them.

Thus, the first important combination is that of lifetimes with *branching ratios*. The measured intensity I_i of a line $k \rightarrow i$ (if it is optically thin, of course)

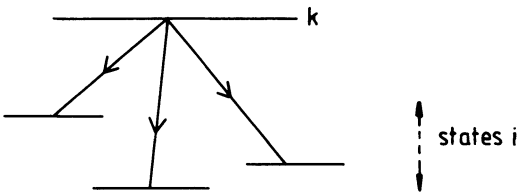

Figure 12.10 Branching ratios. The lifetime of state k is given by $\tau_k = 1/\Sigma_i A'_{ki}$, summed over all states i with which k can combine.

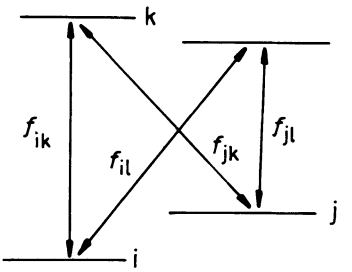

Figure 12.11 Combination of emission and absorption measurements between the four levels k, l, i and j to obtain relative oscillator strengths independent of population.

is given by

$$I_i = C_i A_{ki} h\nu_i N_k$$

where C_i contains not only all geometric factors but also the spectrometer and detector responses at the frequency ν_i. The transition probability of any one line, $k \rightarrow j$, can be written as

$$\frac{A_{kj}}{\sum_i A_{ki}} = \frac{I_j/C_j\nu_j}{\sum_i (I_i/C_i\nu_i)} \Rightarrow A_{kj} = \frac{1}{\tau_k} \frac{I_j/C_j\nu_j}{\sum_i (I_i/C_i\nu_i)}$$

The experimental difficulties of this seemingly simple method are mainly associated with the the Cs. It is not a trivial matter to calibrate detectors and spectrometers accurately over a spectral range that may extend from the infra-red to well into the ultra-violet.

The increasing availability of rather good relative oscillator strengths for the same set of lines from emission, absorption and dispersion methods has led to the exploitation of a technique first used by Ladenburg in the 1930s and now known by various names: 'Ladenburg' and 'bow-tie' among them. It is best explained with reference to Fig. 12.11. Suppose we have two upper levels k and l that can combine with two lower levels i and j by the four transitions shown. From two pairs of absorption (or dispersion) measurements we can find, say,

equivalent widths W:

$$W_{il}/W_{ik}=f_{il}/f_{ik} \quad \text{and} \quad W_{jk}/W_{jl}=f_{jk}/f_{jl}$$

Similarly, from two pairs of emission measurements

$$\frac{I'_{kj}}{I'_{ki}}=\frac{\nu_{kj}A_{kj}}{\nu_{ki}A_{ki}}=\frac{\nu_{kj}^3 f_{jk}}{\nu_{ki}^3 f_{ik}} \quad \text{and} \quad \frac{I'_{lj}}{I'_{li}}=\frac{\nu_{lj}^3 f_{jl}}{\nu_{li}^3 f_{il}}$$

where I' represents the intensity I/C corrected for instrument response, and g-values have been omitted for simplicity. In the first place the consistency of the data can be checked by comparing the emission value of $f_{il}f_{jk}/f_{ik}f_{jl}$ with the same quantity obtained from the Ws. Secondly, if one of the f-values is taken as reference we have four equations to determine three unknowns, so a least-squares best fit can be used to put the four f-values on the same scale, which becomes absolute if the reference value can be made absolute.

The method can be extended to a larger system of transitions so long as each additional level is connected to the network by at least one absorption (or dispersion) and one emission measurement. The larger the number of interlinked levels, the more are the equations overdetermined, and the easier it becomes to identify a bad measurement.

Ideally the method is combined with an absolute lifetime measurement to put all the f-values on an absolute scale. If $k \rightarrow i$ and $k \rightarrow j$ are the only transitions from level k in Fig. 12.11, a measurement of τ_k combined with the branching ratio of these two lines serves to put all the f-values in the network on an absolute scale.

References

DATA COLLECTIONS

1. Wiese, W.L., Smith, M.W. and Glennon, B.M. (1966) Atomic transition probabilities: Vol. I (H to Ne), *NSRDS–NBS 4*. Wiese, W.L., Smith, M.W. and Miles, B.M. (1969) Atomic transition probabilities: Vol. II (Na to Ca), *NSRDS–NBS 22*.
2. Critical compilations by various authors on elements of the Fe group in *J. Phys. Chem. Ref. Data*: **2**, 85 (1973). **4**, 263 (1975). **7**, 495 (1978). **10**, 305 (1981).
   Miles, B.M. and Wiese, W.L. (1969) on Ba I and II, *Atomic Data*, **1**, 1. Martin, G.A. and Wiese, W.L. (1976) on the Li iso-electronic sequence, *J. Phys. Chem. Ref. Data*, **5**, 537.
3. Wiese, W.L. and Martin, G.A. (1980) Wavelengths and transition probabilities for atoms and atomic ions, *NSRDS–NBS 68*; also in (1980) *Handbook of Chemistry and Physics*, CRC Press, Cleveland, Ohio.
4. Fuhr, J.R., Miller, B.J. and Martin, G.A. (1978) Bibliography on atomic transition probabilities 1914–1977, *NBS Spec. Publ. 505*, and Supplement 1, 1977–1980.
5. Huber, K.P. and Herzberg, G. (1979) *Constants of Diatomic Molecules*, Van Nostrand, New York.

REVIEW ARTICLES

6. Huber, M.C.E. and Sandeman, R.J. (1986) The measurement of oscillator strengths, *Rep. Prog. Phys.*, **49**, 397.
7. Imhof, R.E. and Read, F.H. (1977) Measurement of lifetimes of atoms, molecules and ions, *Rep. Prog. Phys.* **40**, 1.
8. Wiese, W.L. (1979) Atomic transition probabilities and lifetimes, in *Progress in Atomic Spectroscopy*, Part B, Plenum Press, New York.

METHODS OF MEASUREMENT

9. Bashkin, S. (ed.) (1976) *Beam Foil Spectroscopy*, Springer-Verlag, Berlin.
10. Berry, H.G. (1977) Beam foil spectroscopy, *Rep. Prog. Phys.*, **40**, 157.
11. Corney, A. (1977) *Atomic and Laser Spectroscopy*, Oxford University Press.
12. Mitchell, A.C.G. and Zemansky, M.S. (1934) *Resonance Radiation and Excited Atoms*, Cambridge University Press (reprinted 1971).
13. *Walther, H. (ed.)* (1976) *Laser Spectroscopy of Atoms and Molecules*, Springer-Verlag, Berlin.

THEORETICAL

14. Cohen, M. and McEachran, R.P. (1980) *Atomic Hartree–Fock Theory* (*Advances in Atomic and Molecular Physics* (D.P. Bates and B. Bederson, eds)), Vol. 16, p. 1.
15. Crossley, R.J.S. (1969) *Calculation of Atomic Transition Probabilities* (*Advances in Atomic and Molecular Physics* (D. P. Bates and B. Bederson, eds)), Vol. 5, p. 237.
16. Garstang, R.H. (1962) Forbidden transitions, *Atomic and Molecular Processes* (D.P. Bates, ed.), Academic Press, New York.
17. Hibbert, A. (1975) Atomic structure calculations, *Rep. Prog. Phys.*, **38**, 1217.

THIRTEEN

Elementary plasma spectroscopy

13.1 Introduction

The preceding chapters have shown that the widths, shapes and intensities of spectral lines depend on the temperature, pressure and electron density of the environment of the atom or molecule, as well as on its intrinsic properties. If the broadening and other physical processes are properly understood and the necessary atomic parameters are known, the spectral lines can give information about the physical conditions in the emitting or absorbing gas or plasma. In laboratory conditions spectroscopic techniques have the advantage over some other methods – those involving probes, for example – that they do not interfere in any way with the plasma, and in the early days of plasma physics spectroscopy was an important diagnostic tool. Other optical diagnostic techniques using lasers and masers have now supplemented and to a considerable extent superseded the purely spectroscopic ones, but there remains an important role for spectroscopy in determining the physical processes going on in the plasma. Moreover, understanding of laboratory plasmas is an important prerequisite to the use of spectroscopic methods on astrophysical plasmas for which no alternative methods are available.

In this chapter a brief account of plasma parameters is followed by the introduction of Saha's equation for ionization equilibrium, a note on excitation and de-excitation by electron collisions, and a discussion of the concepts of temperature and equilibrium conditions. Continuous emission and absorption are then treated, and the remainder of the chapter is concerned with a range of applications.

The word 'plasma' is sometimes used loosely both in this book and elsewhere to describe a mass of hot gas regardless of its degree of ionization. The proper definition of a plasma is quite unambiguous, as shown in the next section, but the word is a convenient shorthand to describe any vapour or gas that is partly ionized.

For the moment, then, we consider a plasma to be an assembly of atoms,

molecules, ions and electrons that may be treated statistically in the same way as a solid, liquid or unionized gas. The plasma as a whole is electrically neutral, the total ionic charge being equal to the number of free electrons. It can be quite easily shown that the charge neutrality must apply to quite a small volume: if N_i and N_e are the number densities of ions (assumed singly charged) and electrons, respectively, a sphere of radius r contains a net charge Q of $\frac{4}{3}\pi r^3(N_i - N_e)e$, and the corresponding rise in potential V is given by

$$V = \frac{Q}{(4\pi\varepsilon_0)r} = \frac{(N_i - N_e)r^2}{3\varepsilon_0}$$

With a typical value of N_e, 10^{22} m^{-3}, and a 1% difference between N_i and N_e the rise in potential at the surface of a sphere of radius 1 mm works out to be nearly a million volts. Such a large potential difference cannot possibly be maintained in a gas, where the particles can move freely to neutralize the excess charge. It is therefore quite safe to take the average particle densities as given by $N_i = N_e$ until one gets down to distances so small that the out-of-balance potential energy is comparable with the thermal energy. The order of magnitude of such a distance is given by the Debye screening radius ρ_D.

NOTE ON UNITS

In the SI system the unit of number density is m^{-3}, but in the references cited in this chapter and in nearly all current work on plasma spectroscopy number densities are given in cm^{-3}. I have used m^{-3} in the equations in this chapter for consistency, but I have deliberately used cm^{-3} descriptively from time to time because the advantage of familiarity with the actual numerical values seems to me to outweigh the disadvantage of having to multiply occasionally by 10^6 when inserting a number density into an equation.

Plasma physicists frequently measure temperature in electron volts. Since $1\text{ eV} = 1.6 \times 10^{-19}$ J, and $kT = 1.4 \times 10^{-23}$ T J, the temperature corresponding to 1 eV is 11 400 K. For semiquantitative purposes 1 eV = 10 000 K is a sufficiently good approximation.

13.2 Debye radius

The distinguishing characteristic of a plasma is the importance of collective effects. Every charged particle interacts with several other particles at the same time because of the long range of electrostatic forces, and the motions of the particles are therefore correlated. The Debye shielding radius ρ_D is a criterion of the importance of collective effects: it measures the distance to which the electric field of an individual ion or electron extends before it is effectively shielded by the oppositely charged particles. Individual interactions between particles are important only over distances less than ρ_D, and over larger distances the

collective effects dominate. The criterion for the existence of a plasma, or the predominance of collective effects, therefore amounts to requiring ρ_D to be much smaller than the dimensions of the assembly of particles.

The expression normally used for the Debye radius is that originally derived by Debye for an electrolyte. Suppose a particle of charge q is injected into an electrically neutral plasma, for which the mean potential is zero and $\bar{N}_i = \bar{N}_e$. In the neighbourhood of q the potential has some value V, and there is a local charge imbalance giving a net charge density $\rho = e(N_i - N_e)$. If the particle density is sufficiently high for V to vary smoothly, unaffected by individual ions or electrons, Poisson's equation must be satisfied:

$$\nabla^2 V = -\frac{\rho}{\varepsilon_0} = -\frac{e(N_i - N_e)}{\varepsilon_0}$$

An electron at potential V has energy $-eV$. In equilibrium the number of electrons with this energy is given by Boltzmann's equation (equation 4.17) in the form

$$N_e = \bar{N}_e \, e^{eV/kT}$$

since $N_e = \bar{N}_e$ when $V = 0$. Similarly for the ions,

$$N_i = \bar{N}_i \, e^{-eV/kT} = \bar{N}_e \, e^{-eV/kT}$$

The differential equation determining V becomes

$$\nabla^2 V = -\frac{e}{\varepsilon_0} \bar{N}_e (e^{-eV/kT} - e^{eV/kT})$$

Assuming that $eV \ll kT$, which will be justified below,

$$\nabla^2 V \simeq \frac{e}{\varepsilon_0} \bar{N}_e \frac{2eV}{kT} = \frac{2e^2 \bar{N}_e}{\varepsilon_0 kT} V$$

For the notional point charge disturbance, V has spherical symmetry, so that

$$\nabla^2 V = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial V}{\partial r} \right)$$

For the boundary conditions $V \to 0$ as $r \to \infty$ and $V \to q/(4\pi\varepsilon_0 r)$ as $r \to 0$, this equation has the solution (easily verified by substitution)

$$V = \frac{q}{(4\pi\varepsilon_0) r} e^{-r/\rho_D} \quad \text{where} \quad \rho_D^2 = \frac{\varepsilon_0 kT}{2e^2 \bar{N}_e}$$

ρ_D determines the 'decay distance' of the potential distribution produced by the charge q, and this is defined as the Debye shielding radius, i.e.

$$\rho_D = \left(\frac{\varepsilon_0 kT}{2e^2 N_e} \right)^{1/2} \tag{13.1a}$$

An expression of the same order of magnitude is obtained if one considers the

penetration of an external field from the boundary of the plasma, in which case ρ_D may be regarded as a skin depth.

If one is concerned with transient plasmas or very rapid disturbances, the ions may not be able to move fast enough to conform to the instantaneous equilibrium charge distribution. Then only the electrons can be effective in the shielding, and ρ_D^2 must be increased by a factor of two:

$$\rho_D = \left(\frac{\varepsilon_0 kT}{e^2 \bar{N}_e}\right)^{1/2} \tag{13.1b}$$

Putting the numerical constants into equation 13.1a gives

$$\rho_D = 50\sqrt{\left(\frac{T}{N_e}\right)} \quad \text{m}$$

where T is in K, $\bar{N}_e$ is in m^{-3}, and ρ_D is in m. For an arc having $T = 10^4$ K, $N_e = 10^{22}\,m^{-3}$ (i.e. $10^{16}\,cm^{-3}$), ρ_D comes to about 5×10^{-8} m, which is obviously very much smaller than the dimensions of the arc column. The high temperature pinches ($T \sim 10^6$ K) of relatively low density ($N_e \sim 10^{14}\,cm^{-3}$), for which $\rho_D \sim 10^{-5}$ m, still have a couple of orders of magnitude in hand. In the solar corona, with $T \sim 10^6$ K and $N_e \sim 10\,cm^{-3}$, ρ_D is of the order of several cm, but of course the dimensions are very many orders of magnitude greater. Figure 13.1 shows values of ρ_D for various types of plasma.

The derivation of ρ_D was subject to the restrictions that V be a smoothly varying function and that $eV \ll kT$. In fact, these two conditions amount to the same, at least so long as the perturbing potential is attributable to an isolated ion or electron. The first requires that the Debye sphere contain a large number of particles:

$$\tfrac{4}{3}\pi\rho_D^3 \bar{N}_e \gg 1 \tag{13.2}$$

Eliminating $\bar{N}_e$ between equations 13.1 and 13.2 gives the inequality

$$\rho_D \gg \frac{3e^2}{(4\pi\varepsilon_0)kT}$$

The mean electrostatic energy $e\bar{V}$ of the electrons surrounding a charge e is of order $e^2/(4\pi\varepsilon_0\rho_D)$, leading directly to the second condition

$$e\bar{V} \simeq \frac{e^2}{(4\pi\varepsilon_0)\rho_D} \ll \frac{kT}{3} \tag{13.3}$$

Plasmas in which the Debye sphere contains only about one particle, so that equation 13.2 is not satisfied, are known as non-Debye plasmas. They exist in stellar interiors and are of considerable theoretical interest. Their formation requires a combination of high ionization with relatively low temperature which may possibly be achievable in the laboratory in laser-compressed plasmas or dense pinch discharges. References [18–20] review the spectroscopy of dense plasmas.

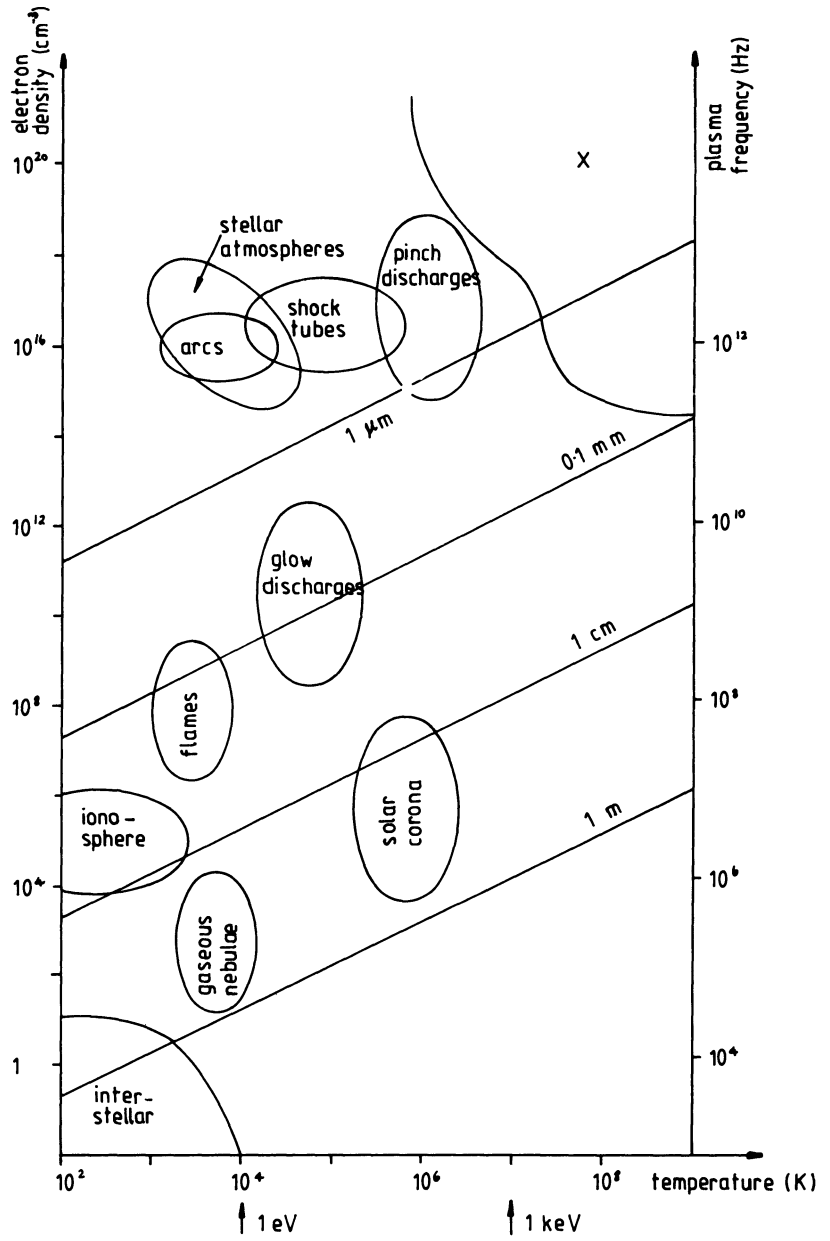

Figure 13.1 Temperature and electron density for different types of plasma. The corresponding plasma frequencies are marked up the right-hand side of the diagram, and the diagonals are lines of constant Debye radius ρ_D. The region marked X corresponds to plasmas generated by lasers, vacuum sparks and plasma focus devices, and to stellar interiors.

13.3 Plasma oscillations

Just as the Debye radius represents a typical length for collective action, so the plasma frequency – or rather, its reciprocal – is a typical time. The plasma frequency is the resonance frequency for collective oscillations of the electrons about their equilibrium positions. It can be derived as follows.

Consider all the electrons in a slab of plasma of unit cross-section and length l to be displaced simultaneously a distance x from their equilibrium positions, where $x \ll l$. The amount of displaced charge is $eN_e x$. The slab now has surface charges $+\sigma$ at one side and $-\sigma$ at the other, with $\sigma = eN_e x$ (see Fig. 13.2). The electric field produced within the slab by this charge is σ/ε_0. Each electron therefore experiences a force $e\sigma/\varepsilon_0$, or $e^2 N_e x/\varepsilon_0$, and its equation of motion, if one neglects collisions and thermal energy, is

$$m\ddot{x} + \frac{e^2 N_e}{\varepsilon_0} x = 0$$

This is the familiar equation of undamped simple harmonic motion. The solution is

$$x = a \sin \omega_p t$$

where the angular plasma frequency ω_p is given by

$$\omega_p^2 = \frac{e^2 N_e}{\varepsilon_0 m} \tag{13.4}$$

The plasma frequency in Hz is

$$\nu_p = \frac{\omega_p}{2\pi} = \left(\frac{e^2 N_e}{(4\pi\varepsilon_0)m}\right)^{1/2} = 9\sqrt{N_e} \text{ Hz} \tag{13.5}$$

where the numerical value is for N_e in m^{-3}. (For N_e in cm^{-3}, $\nu_p = 9000 N_e^{1/2}$). For the laboratory plasmas discussed in the previous section, with N_e in the range 10^{20}–10^{24} m^{-3} (or 10^{14}–10^{18} cm^{-3}), one finds $\nu_p \simeq 10^{11}$–10^{13} Hz, which puts the plasma frequency in the far infra-red, between 3 mm and 30 μm in

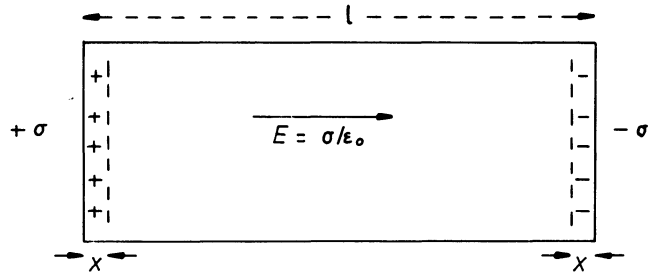

Figure 13.2 Plasma oscillations – see text.

wavelength. Values of ν_p for the various plasmas in Fig. 13.1 are shown on the right-hand side of the diagram.

It should be emphasized that plasma oscillation is not just a resonance oscillation of one electron bound to a nucleus, of the type dealt with in classical absorption and dispersion theory: it is a collective motion of *free* electrons, and the restoring force exists only because all of the electrons are displaced together. This collective behaviour tends to disappear if the electron motions are randomized by collisions, and a necessary condition for its existence is that the time between collisions must exceed the oscillation period: the collision frequency ν_c must be smaller than ν_p.

Equation 13.5 suggests a way of determining N_e, provided that ν_p can be measured. One obvious possibility is to investigate the propagation of radiation through a plasma, since one would expect some sort of change when the frequency of the probe radiation coincides with the plasma frequency. This change is actually a cut-off, as shown below. A preliminary qualitative look is worthwhile for the insight it gives into plasma processes. If we eliminate the electron density between equations 13.1b and 13.4 we have for the relation between Debye length and plasma frequency

$$\rho_D \omega_p = \sqrt{\left(\frac{kT}{m}\right)} \tag{13.6}$$

The mean electron velocity is given by

$$\tfrac{1}{2}m\bar{v}^2 = \tfrac{3}{2}kT$$

so the right-hand side of equation 13.6 is approximately equal to $\bar{v}$. Thus, in the time taken for one plasma oscillation the electron travels about one Debye length. This is just what it needs to do in order to cancel out the disturbance from an external field such as an electromagnetic wave. When $\omega > \omega_p$, the disturbance is too fast for the electrons, but when $\omega < \omega_p$ the electrons can move fast enough to annul the wave and therefore prevent it propagating.

To show this analytically we need an expression for the refractive index of the plasma. Equation 11.25 for the dielectric constant of a medium arising from the oscillations of bound electrons is also valid for free electrons if one sets the resonance frequency ω_0 equal to zero on the grounds that there is no restoring force for *individual* free electrons in a plasma. The complex dielectric constant is then

$$\frac{\varepsilon}{\varepsilon_0} = 1 + \frac{N_e e^2}{\varepsilon_0 m} \frac{1}{-\omega^2 + i\omega\gamma}$$

For negligible damping ($\gamma \simeq 0$) we can forget about the imaginary term and write

$$n_{el}^2 = \frac{\varepsilon}{\varepsilon_0} = 1 - \frac{N_e e^2}{\varepsilon_0 m \omega^2} = 1 - \frac{\omega_p^2}{\omega^2} \tag{13.7}$$

As ω decreases from infinity to ω_p, n_{el} goes from 1 to 0. The phase velocity of the

electromagnetic wave is given in the usual way by $v=c/n$, so it is greater than c when $\omega>\omega_p$, approaches infinity as $\omega\to\omega_p$, and becomes imaginary for $\omega<\omega_p$. In terms of the wave vector k $(=2\pi/\lambda)$ we have

$$v=\frac{\omega}{k}=\frac{c}{(1-\omega_p^2/\omega^2)^{1/2}} \tag{13.8}$$

giving

$$k=\frac{\omega}{c}\left(1-\frac{\omega_p^2}{\omega^2}\right)^{1/2}=\frac{1}{c}\sqrt{(\omega^2-\omega_p^2)} \tag{13.9}$$

As ω decreases, k tends to zero at the plasma frequency and then becomes imaginary. The propagation of electromagnetic energy is determined by the group velocity u, obtained by differentiating equation 13.8:

$$u=\frac{d\omega}{dk}=\frac{c}{\omega}\sqrt{(\omega^2-\omega_p^2)}=c\left(1-\frac{\omega_p^2}{\omega^2}\right)^{1/2} \tag{13.10}$$

u, of course, is always less than c, and it approaches zero at the plasma frequency, where energy propagation ceases. The behaviour of v, u and k as a function of ω is illustrated in Fig. 13.3.

At some distance from the cut-off, when $\omega\gg\omega_p$, and n_{el} is fairly close to unity, equation 13.7 can be expanded to give

$$n_{el}-1=-\frac{\omega_p^2}{2\omega^2}=-\frac{e^2N_e}{2\varepsilon_0 m\omega^2}=-\frac{e^2N_e}{8\pi^2\varepsilon_0 mc^2}\lambda^2 \tag{13.11}$$

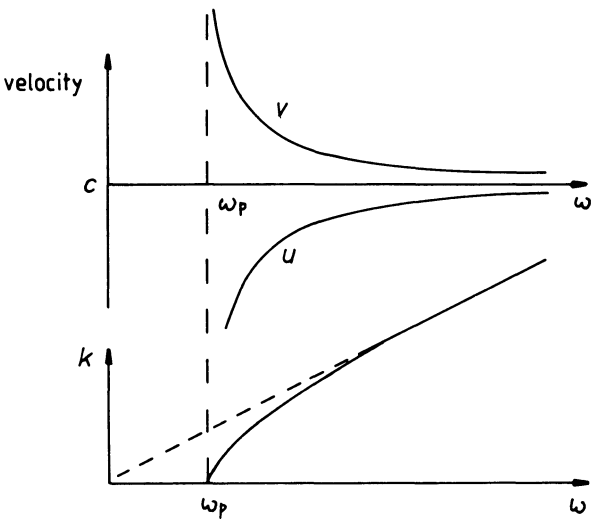

Figure 13.3 Propagation of electromagnetic radiation for plasma frequency ω_p. The diagram shows phase velocity v, group velocity u, and wave vector k as a function of ω.

The comparable expression for bound electrons is equation 11.30, with the damping term negligible so long as we are well away from an actual absorption line. As most strong absorption lines fall in the visible and ultra-violet, it is reasonable to assume $\nu_{ij} \gg \nu$ in the infra-red, in which case equation 11.30 can be expanded as

$$n_{at}-1=\frac{e^2}{8\pi^2\varepsilon_0 m}\sum_j \frac{N_i f_{ij}}{\nu_{ij}^2-\nu^2}=\frac{e^2}{8\pi^2\varepsilon_0 m}\sum_j \frac{N_i f_{ij}}{\nu_{ij}^2}\left(1+\frac{\nu^2}{\nu_{ij}^2}+\cdots\right)$$

$$=\frac{e^2}{8\pi^2\varepsilon_0 mc^2}\sum_j N_i f_{ij}\lambda_{ij}^2\left(1+\frac{\lambda_{ij}^2}{\lambda^2}+\cdots\right) \qquad (13.12)$$

This expression varies very slowly with λ in contrast with the dependence of the electron refractive index in equation 13.11, and it is therefore possible to separate the two contributions to the total refractive index by measuring the latter at two different wavelengths. Such a pair of measurements gives ν_p and hence N_e.

The propagation of electromagnetic waves through a real plasma is considerably more complicated than this simple treatment indicates. Thermal motions and collisions tend to smear out the cut-off frequency, and electrons in highly excited states occupy a position intermediate between free and bound. Moreover, we have not considered at all the effects of external magnetic fields, which are important in many astrophysical and laboratory plasmas. These produce a different type of collective motion, with the electrons spiralling round the field lines with the cyclotron frequency $\omega_c = eB/m$ and the ions with the much lower frequency $(\omega_c)_i = eB/M$.

13.4 Energy distribution and dissociation equilibrium

The expressions describing the distribution of velocity among particles (Maxwell's relation) and the distribution of population among discrete energy levels (Boltzmann's relation) have already been used in this book, the latter on several occasions. We now come to the equations relating the number densities of dissociation products in equilibrium – molecules into atoms, and atoms into ions plus electrons. Like the other distributions, these are obtained by minimizing the free energy of the system, and a text on statistical mechanics or one of the references [1, 2] should be consulted for a derivation of the general relation between number densities and partition functions from which we shall derive the relations for dissociation and ionization equilibrium.

For convenience and also to help elucidate the role of the partition function we rewrite here the Maxwell and Boltzmann relations. Equation 10.15 gave Maxwell's expression for the fraction of particles having a given velocity in the x-direction. Extending this to three dimensions, we have for the fraction of particles with velocity components between v_x and $v_x + dv_x$, v_y and $v_y + dv_y$, v_z

and $v_z + dv_z$:

$$\frac{dN(v_x, v_y, v_z)}{N} = \left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left(-\frac{m(v_x^2 + v_y^2 + v_z^2)}{2kT}\right) dv_x\, dv_y\, dv_z \quad (13.13)$$

If only the total speed matters, not the direction, this can be rewritten as

$$\frac{dN(v)}{N} \equiv f(v)\, dv = \left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left(\frac{mv^2}{2kT}\right) 4\pi v^2\, dv \quad (13.14)$$

The 'volume element' $dv_x dv_y dv_z$ in velocity space has been replaced here by an integration over all angles to give $4\pi v^2\, dv$.

The Boltzmann distribution was given in equation 11.2,

$$\frac{N_j}{N} = \frac{g_j}{Q(T)} \exp\left(-\frac{E_j}{kT}\right) \quad (13.15)$$

in terms of the partition function defined in equation 11.1:

$$Q(T) = \sum_j g_j e^{-E_j/kT} \quad (13.16)$$

Comparing equations 13.14 and 13.15, it can be seen that both have the same exponential $(-E/kT)$ and that $Q(T)$ in the second plays the same role as the normalizing factor $(m/2\pi kT)^{3/2}$ in the first in ensuring that the sum over all possible states adds up to unity. The analogy can be made more obvious by rewriting equation 13.13 in terms of the momenta, $mv_x = p_x$, etc.:

$$\frac{dN(p_x, p_y, p_z)}{N} = \frac{1}{(2\pi mkT)^{3/2}} \exp\left(-\frac{E(v)}{kT}\right) dp_x\, dp_y\, dp_z \quad (13.17)$$

The uncertainty principle does not allow the momentum and the position of a particle to be specified simultaneously with a precision greater than that given by $\Delta p \Delta x = h$. In three dimensions, therefore, $\Delta p_x \Delta p_y \Delta p_z \Delta x \Delta y \Delta z = h^3$. Imagine now a six-dimensional momentum space carved into cells of volume h^3. A particle cannot be specified more precisely than by assigning it to one such cell, but it is quite possible for the particle to have the same energy in any one of several cells, just as a bound electron may be found in several (degenerate) states of the same energy. The statistical weight of the free particle is the number of cells in which it may be found:

$$g = \frac{dp_x dp_y dp_z dx dy dz}{h^3} = \frac{V}{h^3} dp_x\, dp_y\, dp_z$$

where V is the actual volume. With this interpretation of g, equation 13.17 becomes identical with equation 13.15 if the term $1/(2\pi mkT)^{3/2} h^3/V$ corresponds to $1/Q(T)$. This gives for the partition function for translational motion of a free particle:

$$Q_t(T) = \frac{(2\pi mkT)^{3/2}}{h^3} V \quad (13.18)$$

We now come to dissociation and ionization,

$$X+Y \rightleftharpoons Z \quad \text{and} \quad A^+ + e^- \rightleftharpoons A$$

the second being a special case of the first. The relation found from statistical mechanics for the numbers of the three species in equilibrium is

$$\frac{N_X N_Y}{N_Z} = \frac{Q_X^\circ Q_Y^\circ}{Q_Z^\circ} \tag{13.19}$$

where each Q° is the *total* partition function of the species – that is, the product of the internal partition function $Q(T)$ and the free particle translational partition function $Q_t(T)$. For dissociation of the molecule Z into the atoms X and Y equation 13.19 gives

$$\frac{N_X N_Y}{N_Z} = \frac{(2\pi\mu kT)^{3/2}}{h^3} \frac{Q_X Q_Y}{Q_Z} e^{-D/kT} \tag{13.20}$$

The first term on the right comes from the three translational partition functions, using equation 13.18. μ is a sort of reduced mass given by

$$\mu = \frac{M_X M_Y}{M_Z} = \frac{M_X M_Y}{M_X + M_Y}$$

and V has been set equal to unit volume, so that all the Ns are number densities. The Qs are the internal partition functions, the zero of energy being in each case the ground term of the atom or molecule. D is the dissociation energy, and the factor $\exp(-D/kT)$ appears because Q_Z has to be multiplied by $\exp(D/kT)$ to give the state sum of the molecule the same zero of energy as that of the separated atoms.

In many atoms and ions the first excited state is several eV above the ground state, and only the first term in the partition function is significant, at any rate for temperatures up to about 10 000 K. Thus $Q(T) \simeq g_0$ is often a permissible approximation. Moreover, for most simple atoms and ions g_0 is close to unity, so $Q_X Q_Y$ is at most a few units. Conversely, Q_z is very large, because even at moderate temperatures a large number of rotational states, and perhaps a few vibrational states, are occupied; for $T = 1000$ K and a typical value of B (1 cm^{-1}) in equation 11.3a, Q_{rot} is about 700. Consequently an appreciable number of molecules may stay stuck together even when kT is larger than D.

13.5 Saha's equation for ionization equilibrium

Equation 13.19 applied to the process $A^+ + e^- \rightleftharpoons A$ becomes

$$\frac{N_e N_i}{N_a} = \frac{2(2\pi mkT)^{3/2}}{h^3} \frac{Q_i(T)}{Q_a(T)} e^{-\chi/kT} \tag{13.21}$$

where the Ns are again number densities. The first factor on the right is the total partition function for the free electron, and the factor 2 incorporated in it is the state sum for the two possible spin states. The masses of the atom and ion are so nearly equal that their translational partition functions cancel out, leaving only the internal partition functions Q_a and Q_i. χ is the ionization energy, and the factor $\exp(-\chi/kT)$ accompanies $Q_i(T)$ to give the state sums of the ion and the atom the same zero.

Similar equations hold for higher degrees of ionization. The number densities in the zth and $(z+1)$th stages are related by

$$\frac{N_e N_{z+1}}{N_z} = \frac{2(2\pi mkT)^{3/2}}{h^3} \frac{Q_{z+1}(T)}{Q_z(T)} e^{-\chi_z/kT} \tag{13.22}$$

In general, by the time an appreciable number of ions has reached $z+1$ there is a negligible population of $z-1$; it is usually adequate to consider only two stages of ionization for any one species at a given temperature.

The number density N_i refers to the total number of ions per unit volume, regardless of the particular energy levels they may occupy, and the same is true of N_a. The number density for the ground state, or for any particular excited state, is found by combining Boltzmann's and Saha's equations. For example, the number density of neutral atoms in state j is related to the number density of ions in state k (see Fig. 13.4) by

$$\frac{N_e N_{i,k}}{N_{a,j}} = \frac{2(2\pi mkT)^{3/2}}{h^3} \frac{g_{i,k}}{g_{a,j}} e^{-(\chi + E_k - E_j)/kT} \tag{13.23}$$

To get an idea of the degree of ionization in real situations, we shall put some numbers into Saha's equation. Inserting numerical values for the constants,

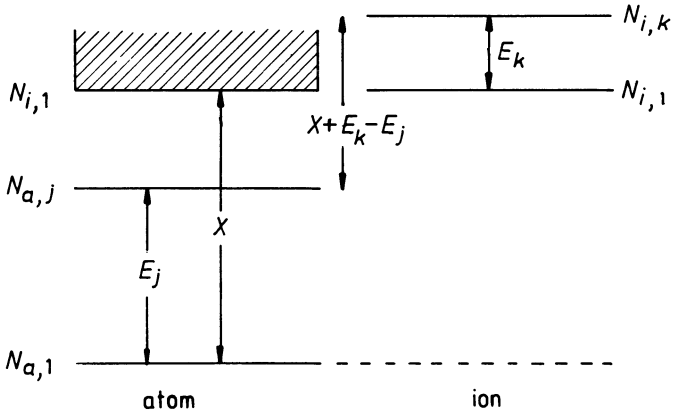

Figure 13.4 Energy levels and population densities for an atom and its ion, as referred to in equation 13.23.

equation 13.21 becomes

$$\frac{N_e N_i}{N_a} = 4.83 \times 10^{21}\, T^{3/2} \frac{Q_i}{Q_a} e^{-1.16 \times 10^4 I/T} \tag{13.24}$$

where the number densities are in m^{-3} and I is the ionization potential in V. Since the internal partition functions of both atoms and ions are likely to be only a few units, as remarked in the previous section, the ratio Q_i/Q_a is usually of order unity and is not very strongly temperature-dependent. There is an appreciable degree of ionization even when $I \gg kT$. With $I = 10$ V (which is mid-range for the periodic table) and $T = 10\,000$ K, the exponential factor is only about 10^{-5}, but the right-hand side of the equation works out to be 4.8×10^{22}. For an electron density of 10^{17} cm^{-3}, or 10^{23} m^{-3}, such as might be found in a high-current arc, we have $N_i \simeq \frac{1}{2} N_a$, or 30% ionization. It is worth contrasting this figure with the population of excited states in the neutral atom: an excited level 10 eV above the ground state at this temperature has a population only about 10^{-5} of that of the ground state.

As a particular example, we calculate the degree of ionization of calcium ($I = 6.1$ V) at this same temperature and electron density. For Ca and Ca^+ the ground states are $^1S_0$ and $^2S_{1/2}$, respectively, giving $Q_i/Q_a = 2$. The ratio N_i/N_a comes to about 100. If only 1% of the calcium atoms are neutral, what about the next stage of ionization? Ca^+ has $I = 11.9$ V, and as Ca^{2+} has a $^1S_0$ ground state the ratio of the partition functions is $\frac{1}{2}$. The Ca^{2+}/Ca^+ ratio is found to be only about 2%; about 97% of the calcium exists as Ca^+.

In a laboratory plasma of relatively low density and high temperature, such as a pinch discharge ($N_e \sim 10^{15}$ cm^{-3}, $T \sim 10^6$ K), equation 13.22 gives $N_{z+1}/N_z \sim 5 \times 10^9 \exp(-I/100)$. Appreciable ionization should be attained for any species having $\exp(-I/100) \sim 10^{-9}$. This corresponds to I being about 2000 V, or say 10–12 stages of ionization. Such high ionization is not in fact attained in most pinch discharges because the electron density is too low and the discharge time too short for the establishment of thermodynamic equilibrium, but in the dense plasma focus device ionization stages up to 18 or so have been reached.

In stellar atmospheres, where the temperature is likely to be a few thousand K, the degree of ionization is higher than in laboratory plasmas of comparable temperature because of the much lower electron density. For example, in the Sun's lower chromosphere $T \sim 7000$ K and $N_e \sim 10^{10}$ cm^{-3}. Astrophysicists usually replace the electron density by the electron pressure, $P_e = N_e kT$, when using Saha's equation, and in this example the appropriate electron pressure is 10^{-9} N m^{-2}. Substituting the calcium values in Saha's equation, one finds that only about 1 Ca atom in 10^7 is neutral in these conditions.

13.6 Depression of ionization potential and Inglis–Teller limit

The distance $\langle r \rangle$ of a valence electron from its parent nucleus increases as n^2, and when $\langle r \rangle$ is approximately equal to ρ_D, the limit of influence of the nucleus,

the electron is no longer effectively bound. Thus 'ionization' takes place when the potential energy of the electron is approximately $-ze^2/(4\pi\varepsilon_0\rho_D)$ instead of zero; ze here is the nuclear charge seen by the valence electron, so $z = 1$ for a neutral atom, 2 for a singly charged ion, etc. To allow for this effect the ionization potential must be reduced by an amount $\Delta\chi$ given by

$$\Delta\chi \simeq \frac{ze^2}{4\pi\varepsilon_0}\sqrt{\left(\frac{e^2N_e}{\varepsilon_0 kT}\right)} = 3\times10^{-11}\,z\sqrt{\left(\frac{N_e}{T}\right)}\ \text{eV}$$

In a typical high current arc, with the parameters given in the last section, $\Delta\chi$ is only about 0.1 eV, which is too small to measure accurately. The experimental uncertainties make it difficult to decide conclusively in favour of any one of the various more sophisticated expressions for $\Delta\chi$.

The depression of the ionization potential removes a difficulty about the divergence of partition functions at high temperatures. In any atom there are an infinite number of bound states crowding up towards the ionization limit, and unless the exponential function $\exp(-E_j/kT)$ is vanishingly small for these states – i.e. $\chi \gg kT$ – the sum diverges; but if the series is terminated at a value of n, n^* say, such that $\langle r_{n^*}\rangle = \rho_D$ the sum is necessarily finite. n^* is related to $\Delta\chi$ by $1/n^{*2} = \Delta\chi$ rydbergs.

This 'depressed series limit' must be distinguished from the superficially similar Inglis–Teller limit. The latter is defined by the n-value at which the individual energy levels are sufficiently Stark-broadened to merge with one another. The dependence of this limit on N_e can be estimated by treating all atoms as hydrogen-like, which is a good approximation at high n-values. The spread of the Stark pattern (Fig. 2.18) is then proportional to $n(n-1)E$, as shown at the end of section 2.9. Averaging over the fields from different neighbours converts the Stark splitting to Stark broadening, proportional at large n to $\langle E\rangle n^2$. However, the separation of states of high n, $\mathrm{d}(1/n^2)/\mathrm{d}n$, is proportional to $1/n^3$. The value of n, say n_m, at which the broadening catches up with the separation defines the Inglis–Teller limit. In the quasi-static approximation $\langle E\rangle$ can be taken as the electric field from a particle at the mean interparticle separation R – i.e. $\langle E\rangle \propto 1/\bar{R}^2$. Since $\bar{R} \propto N_e^{-1/3}$, the broadening should be proportional to $N_e^{2/3}n^2$, and we expect to find $n_m^5 \propto N_e^{-2/3}$.

n_m is in general lower than n^*. This is understandable, in that n_m limits the radius of the atom to the interparticle distance $\bar{R}$, while n^* limits it to ρ_D, which has already been assumed to be greater than $\bar{R}$. The functional dependence of n^* on N_e is also different from that of n_m: from the relations $n^* \propto (\Delta\chi)^{-1/2}$ and $\Delta\chi \propto N_e^{1/2}$, it follows that $n^* \propto N_e^{-1/4}$.

13.7 Electron collisional rates

Excitation and ionization and their inverses, de-excitation and recombination, can be brought about by collisions as well as by absorption and emission of

radiation. This book has hitherto been mainly concerned with the radiative processes, but collisions are a basic ingredient of plasma spectroscopy and play an essential role in the establishment of local thermodynamic equilibrium, as shown in the next section. Whereas the kinetic equilibrium described by Maxwell's distribution is based on elastic collisions, it is the inelastic variety that determine the validity of Boltzmann's and Saha's equations. These inelastic collisions are mentioned at the end of section 10.6 as contributing to atomic linewidths, and as stated there they are predominantly attributable to electrons. The processes of importance are therefore:

$$\text{excitation and de-excitation} \quad A + e^- + \varepsilon \rightleftharpoons A^* + e^-$$

$$\text{ionization and recombination} \quad A + e^- + \varepsilon \rightleftharpoons A^+ + e^- + e^-$$

where ε stands for the kinetic energy transferred to or from the electron.

The cross-sections for these interactions depend both on the transition moment of the atom for the pair of states involved and on the velocity v of the electron. For excitation there must be a threshold velocity v_t defined by the conversion of all the electron kinetic energy into excitation energy:

$$\tfrac{1}{2} m v_t^2 = E_2 - E_1 \tag{13.25}$$

A 'hard sphere' classical argument suggests that for $v \gg v_t$ the cross-section should decrease as $1/v^2$, and it is shown below that the quantum-mechanical Born approximation gives almost the same v-dependence. For neutral atoms the order of magnitude of most excitation cross-sections is around πa_0^2. For transitions in the optical range of 2 or 3 eV and for an electron temperature of 1 eV or so, it is only the electrons in the high energy tail of the Maxwell distribution that satisfy equation 13.25 and are effective in collisional excitation.

Given that we need consider only electrons, we define the cross-sections for collisional excitation and de-excitation between levels 1 and 2 by $\sigma_{12}(v)$ and $\sigma_{21}(v)$, respectively, both being functions of v for the reasons stated. The general form of these cross-sections is sketched in Fig. 13.5(a). The probability per second of an electron of velocity v making a collision is equal to the target area presented to it by a number density N of atoms, each of cross-section σ, in a cylinder of unit area and length v – i.e. $Nv\sigma$. To get the total number of collisions per second per unit volume this must be multiplied by the number density of electrons with velocity v, $dN_e(v)$, and integrated over all v:

$$\begin{aligned} \text{collisions s}^{-1}\,\text{m}^{-3} &= N \int_0^\infty v\sigma(v)\, dN_e(v) = NN_e \int_0^\infty v\sigma(v) f(v)\, dv \\ &= NN_e \langle v\sigma(v) \rangle \end{aligned} \tag{13.26}$$

where $f(v)$ is the electron velocity distribution, given by equation 13.14 if it is Maxwellian and also sketched in Fig. 13.5(a). The quantity $\langle v\sigma(v) \rangle$ is the average value of $v\sigma$ weighted with this velocity distribution and is the area under the curve in Fig. 13.5(b).

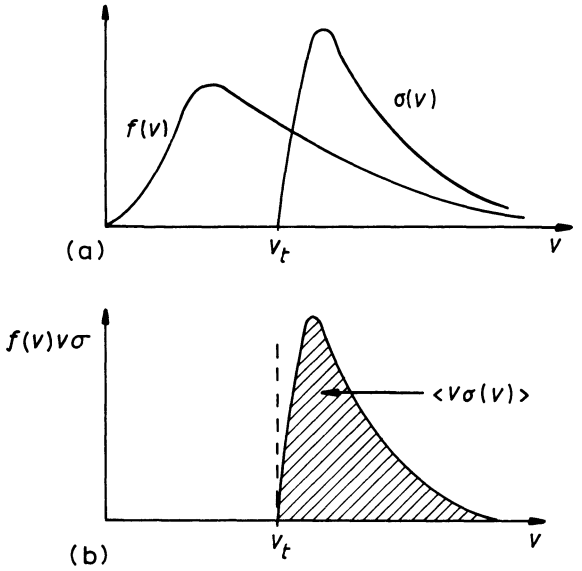

Figure 13.5 (a) Collisional excitation cross-section $\sigma(v)$ as a function of v, together with the Maxwell velocity distribution function $f(v)$. v_t is the threshold velocity. (b) The product $f(v)v\sigma(v)$. The area under the curve is the total collisional probability per atom for unit electron density.

It is convenient to define two collisional coefficients by

$$C_{12} = \langle v\sigma_{12}(v) \rangle \quad \text{and} \quad C_{21} = \langle v\sigma_{21}(v) \rangle \tag{13.27}$$

In terms of these coefficients the transition probabilities per atom are $N_e C_{12}\ \mathrm{s}^{-1}$ upwards and $N_e C_{21}\ \mathrm{s}^{-1}$ downwards.

The calculation of collision cross-sections is beyond the scope of this book, but is treated in a number of standard textbooks (for example [4–6, 21]). Here we give the result of the widely used Born approximation. A free electron travelling in the z-direction is represented by the wave function $\psi = ae^{ikz}$, where $\hbar k$ is equal to the momentum mv. The Born approximation is essentially a first-order perturbation treatment in which this unperturbed wave function is used to evaluate the interaction with the transition moment of the atom. The resulting cross-section for an electric dipole transition is

$$\begin{aligned} \sigma_{12} &= \frac{m^2 e^2}{4\pi\varepsilon_0^2 h^4} \frac{1}{k^2} S_z \ln\left(\frac{2mv^2}{E_2 - E_1}\right) \\ &= \frac{e^2}{16\pi^3 \varepsilon_0^2 h^2} S_z \frac{1}{v^2} \ln\left(\frac{2mv^2}{E_2 - E_1}\right) \end{aligned} \tag{13.28}$$

where S_z is the z-component of the electric dipole line strength of the transition.

Cross-sections for transitions that are optically 'forbidden' – i.e. $S_z = 0$ – depend on the quadrupole term $\langle 2|z^2|1\rangle$; they are smaller, although by not such a large factor as the radiative electric quadrupole transitions, and fall off faster with energy above the threshold. The Born approximation is strictly valid only for electron energies well above threshold, but it often works surprisingly well at lower energies if one accepts that it overestimates the cross-sections near threshold.

Extensive formulae for more accurate calculations and tables of numerical results are given in [21]. For data on collisional ionization see [22].

13.8 Local thermodynamic equilibrium

Reference has been made in several places in this book to equilibrium distributions and to local thermodynamic equilibrium (LTE) as the condition for the validity of the Boltzmann relationship. The time has now come to investigate what is meant by LTE and in what conditions it may be expected to hold.

The four energy distributions that have been used refer to radiative, kinetic, excitation and ionization energy – the Planck, Maxwell, Boltzmann and Saha relations, respectively. Of these, the Planck function and its approximations (Rayleigh–Jeans for long wavelengths and Wien for short wavelengths) are discussed in section 4.2 and the other three are collected in the present chapter as equations 13.14, 13.15 and 13.21. All of them involve a temperature T. If radiation and particles are all bottled up together in an enclosure maintained at constant temperature, there is no doubt about the value to be assigned to T; it is obviously the same for each of the distributions, and the system is said to be in thermodynamic equilibrium. Thermodynamic arguments then lead to the principle of detailed balance, or microscopic reversibility: each energy exchange process must be balanced by its exact inverse. For every photon emitted a photon of the same frequency must be absorbed, for every excitation by electron collision there must be a de-excitation between the same two levels, etc.

In real situations there is often an equilibrium distribution of one or more, but not all, of these forms of energy, and the temperature parameter T may vary from one to another. For example, in a low-pressure discharge tube the gas kinetic temperature may be a couple of orders of magnitude smaller than the Boltzmann temperature, while the spectrum bears no resemblance to a blackbody continuum at any temperature. The form of energy most likely to be out of balance with the others is the radiation energy, since radiative equilibrium requires the plasma to be optically thick at *all* frequencies (cf. section 11.9). In the state known as LTE it is possible to find a common temperature, which may vary from place to place, that fits the Boltzmann and Saha distributions and also the Maxwell distribution for the velocities of the electrons. The criterion for LTE is that collisional processes must be more important than radiative, so that the shortfall of radiative energy does not matter. More precisely, an excited state

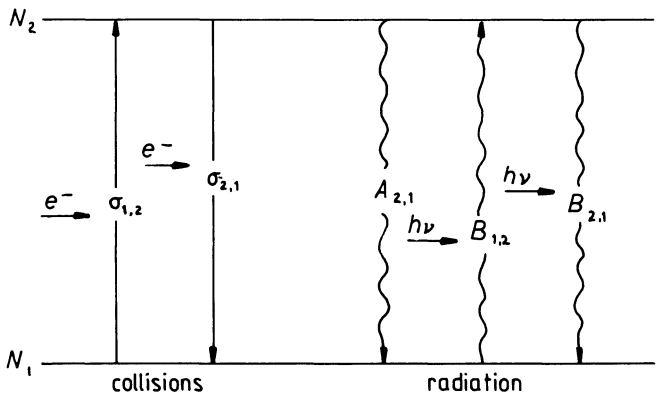

Figure 13.6 Collisional (straight lines) and radiative (wavy lines) processes for a two-level atom.

must have a higher probability of de-excitation by collision than by spontaneous radiation. This is tantamount to requiring a high electron density.

Figure 13.6 shows all the radiative and collisional rates connecting levels 1 and 2. The principle of detailed balance requires that in thermodynamic equilibrium the collisional and radiative rates must balance separately; this principle is used in section 4.5 to derive the relations between the Einstein coefficients. We use it in the same way now to relate C_{12} and C_{21}:

$$N_1 N_e C_{12} = N_2 N_e C_{21}$$

therefore

$$\frac{C_{12}}{C_{21}} = \frac{N_2}{N_1} \Rightarrow \frac{\langle v\sigma_{12}(v)\rangle}{\langle v\sigma_{21}(v)\rangle} = \frac{g_2}{g_1} e^{-(E_2-E_1)/kT} \tag{13.29}$$

where the justification for the last equality is that Boltzmann's relation is valid in thermodynamic equilibrium. The σs depend only on atomic parameters and on v, and so this relation must hold even when the system is not in thermodynamic equilibrium *provided that* the electron velocities still have a Maxwellian distribution at the temperature T that appears on the right-hand side of the equation. The fact that the relation is temperature-dependent means that it is not so general as that connecting the A and B coefficients.

Let us now consider a two-level atom in *any* sort of equilibrium, meaning simply that the populations N_1 and N_2 do not change with time. Equating the total number of transitions per unit time upwards with those downwards, we have

$$N_1\rho(\nu_{12})B_{12} + N_1 N_e C_{12} = N_2\rho(\nu_{12})B_{21} + N_2 A_{21} + N_2 N_e C_{21} \tag{13.30}$$

Obviously if N_e is very small we can recover the relation between the radiative

coefficients given in equation 4.18:

$$N_1 \rho B_{12} = N_2(\rho B_{21} + A_{21})$$

At the other extreme we take ρ to be very small, leaving us with

$$N_1 N_e C_{21} = N_2 N_e C_{21} + N_2 A_{21} \tag{13.31}$$

If N_e is large enough for the last term to be ignored, the population ratio is given by

$$\frac{N_2}{N_1} = \frac{C_{12}}{C_{21}}$$

However, from equation 13.29,

$$\frac{C_{12}}{C_{21}} = \frac{g_2}{g_1} e^{-(E_2 - E_1)/kT} \tag{13.32}$$

Thus, the population ratio is described by the Boltzmann equation with a temperature parameter T identical to that for the electron velocity distribution, exactly as required for LTE. The criterion for its realization is

$$N_e \gg A_{21}/C_{21} \tag{13.33}$$

The frequency and temperature dependence of this inequality can be estimated as follows. From equation 13.28 we have $\sigma \propto S/v^2$, so $C \propto S/v$. Using A_{21} proportional to $v^3 S$ (Table 11.1) equation 13.33 can be written as

$$N_e \gg \text{const.}\, v v^3 = \text{const.} \sqrt{(T)}\, (\Delta E)^3$$

A more thorough treatment leads to the numerical form (see, for example [11]):

$$N_e \gg 1.6 \times 10^{18} \sqrt{(T)}\, (\Delta E)^3 \quad \text{m}^{-3} \tag{13.34}$$

where T is the electron temperature in K and ΔE is the energy difference in eV between the state in question and any neighbouring state to which it can make a transition.

This criterion is most difficult to satisfy for low-lying states where ΔE is large. At 10 000 K N_e must be greater than 10^{16}–10^{17} cm^{-3}. In a free-burning arc at atmospheric pressure running at a few ampères, N_e is at least an order of magnitude too small for LTE to exist except in rather highly excited states. It is for this reason that stabilized arcs capable of running at high currents have been developed for use in experiments requiring knowledge of populations.

The temperature-dependence is actually greater than would appear from equation 13.34 because at high temperature one is probably concerned with ions rather than neutral atoms, and ΔE scales as z^2, where z is the degree of ionization. The critical value of N_e therefore increases as z^6. Nevertheless, for any N_e and any z it is possible at high enough excitation to reach a limit where the states are crowded closely enough together for equation 13.34 to hold; this is known as the 'thermal limit' for the particular values of N_e, T and z, and the plasma is said to be in partial LTE.

Many high-temperature laboratory plasmas are transient, lasting 1 μs or less. In such cases it is necessary to estimate whether there is sufficient time for the establishment of collisional equilibrium. Particles accelerated by strong fields must have a large number of collisions to randomize their velocities before it is valid to speak of a temperature, and the time required to establish LTE varies inversely with N_e. A full discussion may be found in references such as [9, 11, 12]. To give an example, full LTE in a plasma at 10 000 K would take about 1 μs for $N_e \sim 10^{16}$ cm^{-3} if collisional processes only were involved, but the time may be significantly shortened by radiative excitation. The time required to establish partial LTE, among the higher states only, in this example is about 10 ns.

Another point for consideration is whether the gas kinetic temperature determining the translational velocities of the ions is equal to the electron temperature. Elastic collisions are very inefficient in transferring energy from electrons to heavy particles because of the large mass difference. As only about $2m/M$ of energy is transferred at each collision, about 1000 collisions are required for approximate equipartition of energy. Nevertheless, in the example given above the time for kinetic equilibrium between electrons and hydrogen ions is only about 10 ns, and the transfer from H-ions to other ions and neutrals is an efficient one.

13.9 Coronal equilibrium and ionization equilibrium

The importance of LTE is that relative populations can be found without any knowledge of cross-sections and transition probabilities. In other forms of equilibrium some at least of these quantities have to be known. The most important form of equilibrium after LTE is known as coronal equilibrium because it is applicable to the Sun's corona where temperature is high ($\sim 10^6$ K) and electron density low ($\sim 10^8$ cm^{-3}). The radiation density is also low, and it is assumed that upward transitions are due to electron collisions while downward transitions occur by spontaneous radiation. The two-level atom of Fig. 13.6 is not a very realistic model, but it will serve to show the considerations involved. For $A_{21} \gg N_e C_{21}$, equation 13.31 becomes

$$N_1 N_e C_{12} = N_2 A_{21}$$

Therefore

$$\frac{N_2}{N_1} = N_e \frac{\langle v\sigma_{12}(v)\rangle}{A_{21}} \tag{13.35}$$

Thus, the population ratio depends on knowing N_e and two rate coefficients.

The inequality assumed above together with equation 13.29 shows that the population ratio is below the Boltzmann value:

$$\frac{N_2}{N_1} \ll \frac{g_2}{g_1} e^{-(E_2 - E_1)/kT}$$

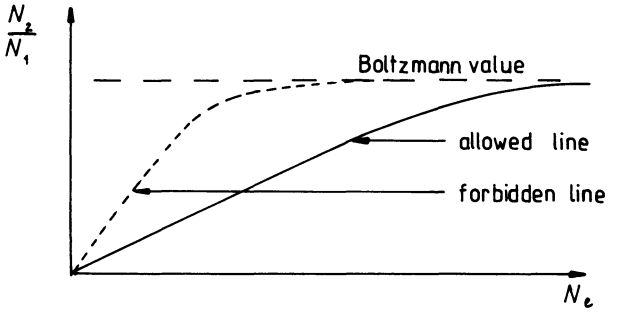

Figure 13.7 Population ratio N_2/N_1 for a two-level atom as a function of N_e when the two levels are connected by an allowed transition (solid curve) or a 'forbidden' transition (broken curve).

This is seen rather clearly from the plot of N_2/N_1 against N_e in Fig. 13.7. The ratio increases linearly at first with N_e, following equation 13.35, but at large N_e it approaches the constant LTE value. For forbidden lines, having small A-values, the LTE condition is reached at lower electron densities. (Note that the ratio of A_{21} to C_{21} in equation 13.33 does not have the same value for forbidden as for allowed transitions.) The appearance of forbidden lines is, indeed, one of the features of the low-density plasmas which can be described by coronal equilibrium. Although excitation rates are low, the only route down once the atom or ion is excited is by spontaneous radiation.

The arguments used to show that LTE is more easily established among higher excited states show also that coronal equilibrium is likely to persist for the lower states when the electron density is high enough to invalidate it for the upper states. It is quite possible to have the populations of the higher states following LTE while those of the lower states follow coronal equilibrium.

Rather similar arguments apply to ionization and recombination. If the three-body recombination rate

$$A^+ + e^- + e^- \rightarrow A + e^- + \varepsilon$$

exceeds the radiative recombination rate

$$A^+ + e^- \rightarrow A + h\nu$$

the ionization ratio is described by Saha's equation. There is also an astrophysically important category of plasmas, stellar atmospheres, in which photoionization exceeds collisional ionization, and Saha's equation is valid on the basis of the balance of radiative transitions only. However, in coronal conditions we find, as for excitation, collisional ionization balancing radiative recombination, and we are back to having to know the individual cross-sections. Both the upward rate

$$A + e^- + \varepsilon \rightarrow A^+ + e^- + e^-$$

and the downward rate

$$A^+ + e^- \rightarrow A + h\nu$$

depend on electron density, and the ratio N_i/N_a is apparently independent of N_e (in contrast to equation 13.35) and is a function of T only. Actually the two-level model is entirely inappropriate to describe the equilibrium; ionization is predominantly from the more populated lower levels, whereas recombination can take place into any of the bound states. A full discussion is given in Mihalas' book [3].

13.10 Continuous emission and absorption

Local thermodynamic equilibrium plasmas would be of singularly little spectroscopic interest if there were not appreciable absorption and emission of radiation in spite of the dominant role played by electron collisions in determining population ratios. The same is true of ionization and recombination: although the collisional processes are nearly always important in determining the degree of ionization, it is the radiative processes of photo-ionization and radiative recombination that are of spectroscopic interest. These are represented by

$$A_{(j)} + h\nu \rightleftharpoons A^+ + e^- + \varepsilon$$

where

$$h\nu = \chi - E_j + \varepsilon \tag{13.36}$$

and ε ($= \frac{1}{2} mv^2$) is the kinetic energy of the free electron (Fig. 13.8). The positive energy states above the ionization limit are not quantized, so the radiation, in both emission and absorption, is a continuum extending in the short-wave-

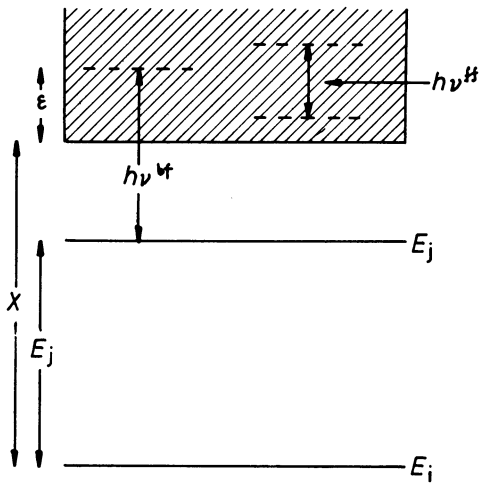

Figure 13.8 Bound–free and free–free transitions – see text.

length direction from the line series limit corresponding to the particular lower state E_j. This so-called bound-free (bf) continuum is characterized by steps, or edges, as shown in Fig. 13.10, below, corresponding to the different series limits; as ν is increased, a new chunk of continuum is added whenever $h\nu$ reaches a new bound level.

Free–free transitions (ff), also illustrated in Fig. 13.8, are associated with loss or gain of energy by an electron in the field of an ion. This type of radiation is expected classically from a charged particle constrained to follow a curved path because the particle is necessarily accelerated in the process. It is otherwise known as 'Bremsstrahlung' (literally, braking radiation). By energy conservation $h\nu = \Delta\varepsilon$, and there are no characteristic edges. At long wavelengths the free–free emission approaches the blackbody function.

Both bound–free and free–free transitions are important in both laboratory and astrophysical plasmas. The principal features are discussed in this and the next two sections, and references such as [1–3, 12] should be consulted for a fuller treatment.

For bound–free radiation there must exist a general relation between the cross-sections for photo-ionization and radiative recombination analogous to that between the Einstein coefficients for line radiation. The relation can be obtained by assuming the system to be in LTE, so that the source function is equal to the blackbody function as in equation 11.50:

$$\frac{j_\nu}{k'_\nu} = L^{\mathrm{B}}(\nu, T) \tag{13.37}$$

k'_ν here is the absorption coefficient corrected for stimulated emission (equation 11.19). It is convenient to work with cross-sections rather than absorption and emission coefficients, so we define the continuous absorption cross-section in the same way as for a line (equation 11.18):

$$\alpha_j(\nu) = k_\nu / N_j \tag{13.38}$$

$\alpha(\nu)$ can be corrected for stimulated emission in the same way as k_ν for line radiation – that is, in LTE

$$\alpha'_j(\nu) = \alpha_j(\nu)\,(1 - \mathrm{e}^{-h\nu/kT}) \tag{13.39}$$

The cross-section for radiative recombination depends on the electron velocity v as well as the particular bound level j and is designated by $\sigma_j(v)$ (Fig. 13.9). The rate of capture of electrons of velocity v to the state j is then $\mathrm{d}N_{\mathrm{e}}(v)v\,\sigma_j(v)$ per ion. For each recapture the energy radiated per unit solid angle is $h\nu/4\pi$. The volume emissivity for a number density N_{i} of ions is given by

$$j_\nu \mathrm{d}\nu = \frac{h\nu}{4\pi} N_{\mathrm{i}}\, v\, \sigma_j(v)\, \mathrm{d}N_{\mathrm{e}}(v)$$

where $\mathrm{d}\nu$ is the frequency spread corresponding to the velocity spread $\mathrm{d}v$. This is

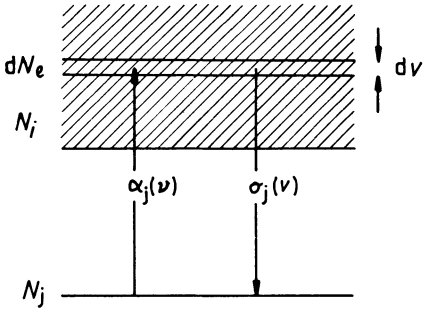

Figure 13.9 Photo-ionization and radiative recombination. N_i is the total number density of ions and dN_e is the number density of electrons with velocities in the range v to $v+dv$.

found by putting $\varepsilon = \frac{1}{2}mv^2$ in equation 13.36 and differentiating:

$$h\,dv = mv\,dv$$

With this substitution and Maxwell's relation for dN_e, we have for LTE

$$j_\nu = \frac{h^2}{m}\left(\frac{m}{2\pi kT}\right)^{3/2} N_i N_e\, v\, v^2\, \sigma_j(v) e^{-\varepsilon/kT} \tag{13.40}$$

and

$$\frac{j_\nu}{k'_\nu} = \frac{N_i N_e}{N_j}\frac{h^2}{m}\left(\frac{m}{2\pi kT}\right)^{3/2} v\, v^2 \frac{\sigma_j(v)}{\alpha_j(v)} \frac{e^{-\varepsilon/kT}}{1-e^{-h\nu/kT}} = L^B(\nu, T) \tag{13.41}$$

In LTE the term $N_i N_e/N_j$ can be evaluated from Saha's equation in the form of equation 13.23. We assume the ion to be in its ground state with statistical weight g_i. After substituting the Planck function (equation 4.12) for L^B and generally tidying up, the equation becomes

$$\frac{\sigma_j(v)}{\alpha_j(v)} = \frac{h^2}{m^2c^2}\frac{v^2}{v^2}\frac{g_j}{g_i} e^{-(\chi - E_j + \varepsilon)/kT} \frac{e^{h\nu/kT}-1}{1-e^{-h\nu/kT}}$$

The last term on the right is simply $e^{-h\nu/kT}$, and the exponentials conveniently vanish completely by virtue of the relation of equation 13.36, leaving

$$\frac{\sigma_j(v)}{\alpha_j(v)} = \frac{h^2}{m^2c^2}\frac{g_j}{g_i}\frac{v^2}{v^2} \tag{13.42}$$

This is known as the Milne or Einstein–Milne relation. Although it was derived for a system in LTE we can use the usual arguments, that the cross-sections are intrinsic atomic constants, to take the relation as valid in any conditions. One of its practical uses is that capture cross-sections, and hence emissivities, can be found from the more easily measured photo-ionization cross-sections.

13.11 The continuous absorption cross-section

The photo-ionization cross-section may be calculated in the same way as the cross-section for a bound–bound transition if the relevant atomic and ionic wave functions are known. In practice such calculations are difficult, and they are also somewhat unreliable except for hydrogen and the H-like ions. The most widely used expression for $\alpha(\nu)$ is based on a semi-classical calculation that extends the cross-sections for bound–bound H-like transitions into the realm of imaginary quantum numbers beyond the series limit [1, 3]. For absorption from the level of principal quantum number n the result is

$$\alpha_n(\nu) = \frac{64}{3\sqrt{3}} \pi a_0^2 \alpha \left(\frac{\chi_H}{h\nu}\right)^3 \frac{Z}{n^5} G_n^{bf}(\nu, Z) \tag{13.43}$$

where a_0 is the first Bohr radius, α is the fine-structure constant ($e^2/4\pi\varepsilon_0 hc$), χ_H is the ionization energy of neutral hydrogen and Z is the effective nuclear charge (1 for a neutral atom, 2 for a single ion, etc.). $G_n^{bf}(\nu, Z)$ is the Gaunt correction factor, introduced to take care of the discrepancy between the semi-classical and quantum-mechanical calculations. It is a function of n, ν and Z that is close to one for H-like ions. As all atoms and ions become increasingly H-like towards the more highly excited states, equation 13.43 is actually a reasonably good approximation for the higher states of any species. The detailed calculations that have been carried out for a number of atoms and ions (mainly those of astrophysical importance) show departures from H-like behaviour in the lower states, even in the alkali-like spectra, in which the continuous absorption passes through a minimum beyond the series limit instead of decreasing steadily with $1/\nu^3$.

Insofar as equation 13.43 can be taken as an acceptable estimate for the cross-section for atoms in state n, the total bound–free absorption coefficient at frequency ν is found by multiplying by N_n and summing over all bound levels n that lie within a distance $h\nu$ of the ionization limit. For the first step we have to assume Boltzmann's relation to hold in order to get anywhere at all, remembering that only partial LTE is required so long as $h\nu$ is much smaller than the ionization energy so that the lower states are not involved. For the H-like case g_n in equation 13.15 is $2n^2$ and E_n can be put in the form $Z^2\chi_H(1 - 1/n^2)$, giving

$$N_n = \frac{N_a}{Q_a(T)} 2n^2 \exp\left[-\frac{Z^2\chi_H}{kT}\left(1 - \frac{1}{n^2}\right)\right]$$

N_a and $Q_a(T)$ here are the number density and partition function of the absorbing atom or ion. To avoid the very cumbersome notation necessary to designate the photo-ionization of an ion of charge z to one of charge $z + 1$, we shall simply call the initial state 'the atom' and the final state 'the ion'. The cut-off for the sum can be expressed in terms of a quantum number n_c determined by

$$n_c^2 \geqslant \frac{Z^2\chi_H}{h\nu} \tag{13.44}$$

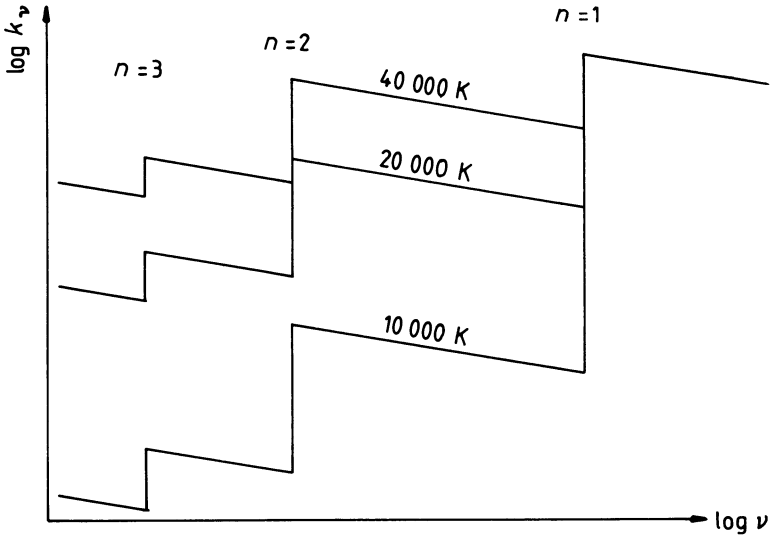

Figure 13.10 Bound–free absorption coefficient as a function of frequency for a H-like atom.

The final expression is then

$$k_\nu^{\mathrm{bf}} = \frac{64}{3\sqrt{3}} N_a \pi a_0^2 \frac{\alpha}{h^3} \frac{\chi_H^3}{Q_a(T)} \frac{2Z^4}{\nu^3} e^{-Z^2\chi_H/kT} \sum_{n \geqslant n_c} \frac{1}{n^3} e^{Z^2\chi_H/n^2kT} G_n \quad (13.45)$$

A plot of log k_ν against log ν has the sawtooth form shown in Fig. 13.10. Whenever ν gets large enough to reach a new bound level, the absorption rises sharply. It then falls off as $1/\nu^3$ until the next bound level is reached. The physical reason for the strong temperature dependence illustrated in the figure is that at low temperatures only the lower levels are sufficiently populated to contribute significantly to photo-ionization; k_ν is thus very small except at short wavelengths. At all temperatures the edges get less pronounced as n increases because the population increments become smaller. For this reason the sum in equation 13.45 can be replaced by an integral for the higher levels, $n \geqslant n'$, where n' is conventionally taken as 5. If we set the Gaunt factor equal to unity and write $(Z^2\chi_H/kT) = \Phi$, this integral becomes

$$\sum_{n=n'}^{\infty} \frac{1}{n^3} e^{\Phi/n^2} \Rightarrow \int_{n'}^{\infty} \frac{1}{n^3} e^{\Phi/n^2} \, dn = -\frac{1}{2} \int_{n'}^{\infty} e^{\Phi/n^2} \, d\left(\frac{1}{n^2}\right)$$

$$= -\frac{1}{2\Phi} [e^{\Phi/n^2}]_{n'}^{\infty} = \frac{1}{2\Phi} (e^{\Phi/n'^2} - 1)$$

From this we get a convenient long wavelength approximation for k_ν . If n_c is

above n', the condition for which is

$$h\nu < \frac{Z^2\chi_H}{n'^2} \quad \text{or} \quad \lambda > \frac{2.5}{Z^2}\,\mu\text{m}$$

the sum term disappears, leaving only the integral with n_c as its lower limit:

$$k_\nu^{bf} = \frac{64}{3\sqrt{3}} N_a \pi a_0^2 \frac{\alpha k}{h^3} \frac{\chi_H^2}{Q_a(T)} \frac{Z^2}{\nu^3} T e^{-Z^2\chi_H/kT} (e^{h\nu/kT} - 1) \qquad (13.46a)$$

(The last exponential is obtained by using equation 13.44 for n_c.) If a_0, α and χ_H are expressed in terms of e, m, h and c, this becomes

$$k_\nu^{bf} = \frac{16\pi^2}{3\sqrt{3}} \frac{k}{ch^4} \left(\frac{e^2}{4\pi\varepsilon_0}\right)^3 \frac{2N_a}{Q_a(T)} \frac{Z^2}{\nu^3} T e^{-Z^2\chi_H/kT} (e^{h\nu/kT} - 1) \qquad (13.46b)$$

Free–free transitions, which have not so far been considered, become important at high temperature and long wavelength. A semi-classical expression for the free–free cross-section of a H-like ion can be obtained by the same device of imaginary quantum numbers as was used for the bound–free cross-section [1, 3]. The result for an electron with velocity between v and $v + dv$ is

$$\alpha_v(v)\,dv = \frac{4\pi}{3\sqrt{3}} \left(\frac{e^2}{4\pi\varepsilon_0}\right)^3 \frac{1}{m^2c^2h} \frac{Z^2}{vv^3} G^{ff}(v, Z, T)\,dv \qquad (13.47)$$

where $G^{ff}(v, Z, T)$ is another Gaunt correction factor, again close to unity so long as kT is below the ionization energy $Z^2\chi_H$. To turn the cross-section into an absorption coefficient, $\alpha_v(v)$ must first be multiplied by the number density of ion–electron pairs, $N_i\,dN_e(v)$ and then integrated over all v for a fixed ν. On the assumption of a Maxwellian electron velocity distribution,

$$\begin{aligned} k_\nu^{ff} &= N_i N_e \frac{16\pi^2}{3\sqrt{3}} \frac{1}{ch(2\pi mk)^{3/2}} \left(\frac{e^2}{4\pi\varepsilon_0}\right)^3 \frac{k}{\sqrt{T}} \frac{Z^2}{\nu^3} G^{ff} \\ &= \frac{16\pi^2}{3\sqrt{3}} \frac{k}{ch^4} \left(\frac{e^2}{4\pi\varepsilon_0}\right)^3 \frac{Z^2}{\nu^3} \frac{2N_a}{Q_a(T)} T e^{-Z^2\chi_H/kT} G^{ff} \end{aligned} \qquad (13.48)$$

The last line is obtained by using Saha's equation with Q_i equal to unity because a H-like ion is just a bare nucleus.

k_ν^{bf} and k_ν^{ff} have the same $1/\nu^3$-dependence. At long wavelengths, when the edge structure can be ignored, k_ν^{ff} in equation 13.48 is actually identical with k_ν^{bf} in equation 13.46 except for the last term and the Gaunt factor G^{ff}. If this is set equal to unity, the combined absorption coefficient assumes the relatively simple form

$$k_\nu^{cont} = k_\nu^{bf} + k_\nu^{ff} \Rightarrow \frac{16\pi^2}{3\sqrt{3}} \frac{k}{ch^4} \left(\frac{e^2}{4\pi\varepsilon_0}\right)^3 \frac{Z^2}{\nu^3} \frac{2N_a}{Q_a(T)} T e^{-(Z^2\chi_H - h\nu)/kT} \qquad (13.49)$$

It must be remembered that this expression is valid only for partial LTE and

that it is a long wavelength approximation, ignoring the edge structure in the visible and ultra-violet.

13.12 The continuous emission

The emission counterpart to k_ν^{bf} in equation 13.46 is j_ν^{bf}, the spectral radiance per unit volume at frequency ν for free–bound radiative recombination, integrated over all electron velocities and summed over all bound levels within a range $h\nu$ below the ionization limit. The easiest way to derive it is to start from equation 13.45, convert k_ν into k' via equation 11.19 or 13.39, find j_ν from the source function relation of equation 13.37, and finally use Saha's equation to express it in terms of the number densities of the initial state particles, ions and electrons, rather than that of neutral atoms. This route gives a misleading impression of dependence on LTE. The alternative is to start from the semi-classical calculation of $\alpha_n(\nu)$ given in equation 13.43 and use the Milne relation (equation 13.42), which is independent of LTE, to find the radiative capture cross-section to state n, $\sigma_n(\nu)$. Equation 13.40 converts $\sigma_n(\nu)$ to j_ν for state n, assuming a Maxwellian velocity distribution for the electrons. It remains to sum over all states n as far as the cut-off n_c given by equation 13.44,

$$n_c \geqslant \frac{Z^2\chi_H}{h\nu}$$

to reach the same end-point as the first route:

$$j_\nu^{\text{bf}} = \frac{128\pi^4}{3\sqrt{3}}\frac{m}{c^3h^2(2\pi mk)^{3/2}}\left(\frac{e^2}{4\pi\varepsilon_0}\right)^5 N_e N_i \frac{Z^4}{T^{3/2}} e^{-h\nu/kT} \sum_{n \geqslant n_c} \frac{1}{n^3} e^{Z^2\chi_H/n^2kT} G_n^{\text{bf}} \tag{13.50}$$

The *only* equilibrium assumption made is that the electron velocities obey Maxwell's relation.

Evidently equation 13.50 has the same edge structure as equation 13.45, but the emission falls off exponentially between the edges instead of as $1/\nu^3$. This can be seen on the right of Fig. 13.11. At wavelengths long enough to ignore the edge structure we have as the counterpart to equation 13.46:

$$j_\nu^{\text{bf}} = \frac{16\pi}{3\sqrt{3}}\frac{1}{mc^3(2\pi mk)^{1/2}}\left(\frac{e^2}{4\pi\varepsilon_0}\right)^3 N_e N_i \frac{Z^2}{\sqrt{T}}(1 - e^{-h\nu/kT}) \tag{13.51}$$

where, as before, G has been set equal to unity.

The free–free emission can be derived in a similar fashion from either equation 13.47 or equation 13.48. Again ignoring the Gaunt correction one obtains

$$j_\nu^{\text{ff}} = \frac{16\pi}{3\sqrt{3}}\frac{1}{mc^3(2\pi mk)^{1/2}}\left(\frac{e^2}{4\pi\varepsilon_0}\right)^3 N_e N_i \frac{Z^2}{\sqrt{T}} e^{-h\nu/kT} \tag{13.52}$$

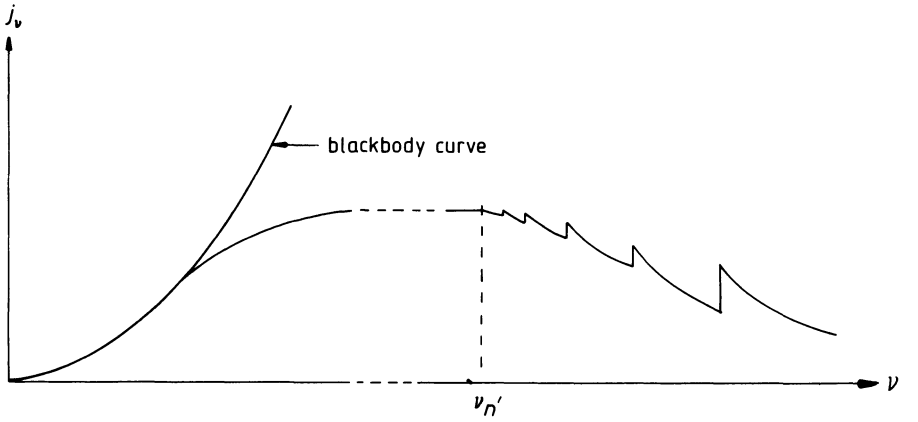

Figure 13.11 Continuous emission coefficient as a function of frequency. $\nu_{n'}$ is the frequency above which edge structure is significant.

on condition only that the electrons obey Maxwell's equation. As with the absorption, we can combine the two types of emission to find the total radiance at long wavelengths:

$$j_\nu^{\text{cont}} = j_\nu^{\text{bf}} + j_\nu^{\text{ff}} = \frac{16\pi}{3\sqrt{3}} \frac{1}{mc^3(2\pi mk)^{1/2}} \left(\frac{e^2}{4\pi\varepsilon_0}\right)^3 N_e N_i \frac{Z^2}{\sqrt{T}} \qquad (13.53)$$

This expression has the interesting property of being independent of frequency – that is, the total continuous emission per unit frequency interval is constant for $\nu < \nu_{n'}$, the frequency at which the edge structure becomes significant. This relation does not depend on the existence of LTE, the only requirement being that the electrons follow a Maxwellian distribution.

At very small values of ν, as $h\nu/kT \to 0$, j_ν is increasingly dominated by the free–free term. However, j_ν cannot continue constant to indefinitely long wavelengths because, as shown in Fig. 13.11, it must eventually hit the blackbody curve, represented in this region by the Rayleigh–Jeans approximation of equation 4.14 – in other words, the plasma becomes optically thick. Thereafter the radiance follows the blackbody curve, falling off as ν^2. The reason for the increasing optical depth is clear from equation 13.49: for $h\nu/kT \to 0$, the absorption coefficient increases as $1/\nu^3$ and must eventually become very large, whatever the particle density. As a matter of fact, it is quite possible for the plasma frequency to fall somewhere in the region where the gas is still optically thin, and in that case the emission falls abruptly to zero at $\nu = \nu_p$. This behaviour is discussed further in section 13.16.

Energy losses through Bremsstrahlung are a significant cause of cooling in laboratory plasmas. The Z^2-factor in equation 13.52 shows that it is important

to eliminate as far as possible impurities of high atomic number from a hydrogen or helium plasma if the object is to get it hot.

13.13 Other continua

The equations for continuous emission and absorption given in the last two sections are strictly applicable only to H-like ions. The general behaviour for many-electron ions is similar, but significant deviations are to be expected at shorter wavelengths when low-lying states are involved in the free–bound transitions. An additional complication may arise from auto-ionization, which causes anomalous bumps, or resonances, in the photo-ionization cross-section (section 2.10). The inverse process of dielectronic recombination (capture of an electron into a bound state above the ionization limit) may enhance the electron capture cross-section sufficiently to have a significant effect on the calculation of rate processes in plasmas.

The negative hydrogen ion, a proton with two electrons, is an important constituent of many stellar atmospheres [1–3], and other negative ions play significant roles in some laboratory plasmas. Negative ions have only one stable state, and their emission and absorption is therefore all continuous. Photo-ionization leaves the neutral H-atom in one of its bound states, the extra energy being carried away by the electron:

$$H^- + h\nu \rightarrow H + e^- + \varepsilon \quad \text{with} \quad h\nu = \chi^- + E_n + \varepsilon$$

χ^- is the detachment energy, or electron affinity. For H^-, χ^- is 0.75 eV, which gives 1.65 μm for the photo-ionization threshold. As the wavelength is decreased below this value, the absorption cross-section rises to a peak at about 850 nm and then drops again.

Photo-detachment cross-sections are of the same order as photo-ionization cross-sections, and the importance of negative ions therefore depends on their relative number densities. Saha's equation is applicable in the form

$$\frac{N_a N_e}{N^-} = \frac{(2\pi mkT)^{3/2}}{h^3} \frac{2Q_a(T)}{Q^-(T)} e^{-\chi^-/kT}$$

Not surprisingly, the formation of negative ions is favoured by large N_e and small T. References [7, 8] give detailed information.

Another possible source of Bremsstrahlung is the cyclotron radiation from electrons moving in magnetic rather than ionic fields. The best known example is the electron synchrotron described as a spectroscopic source in section 4.9.3. In pinch discharges cyclotron radiation in the magnetic field containing the plasma may make a significant contribution to the free–free radiation. Its importance increases as T^2, in contrast with the $1/T$ dependence of thermal Bremsstrahlung. Cyclotron radiation is important in the many astrophysical

plasmas where there are strong magnetic fields, and it is held responsible for the strong continuous emission from certain supernovae.

13.14 Applications of plasma spectroscopy

Plasma spectroscopy developed initially as a method of determining number densities and temperature parameters in plasmas. With the development of other diagnostic techniques – notably laser scattering and laser interferometry – much of the interest in laboratory plasma spectroscopy has switched into three different directions. First, a plasma may be regarded simply as a light source for the spectra of highly ionized species or for the determination of oscillator strengths and collision cross-sections, with T and N_e determined by other methods. Secondly, in a plasma of known composition and temperature the interactions of the particles with each other, with the plasma fields and with the radiation can be investigated through the broadening and shifts of spectral lines, the appearance of forbidden lines and satellites, the relaxation rates of the plasma after excitation or following a perturbation by a laser pulse, and so on. These aspects of plasma spectroscopy may be followed up in references [11–20], which themselves give many further references. Thirdly, in very dense laser-generated plasmas the main non-spectroscopic diagnostics (refractive index measurements, Thompson scattering, etc.) tend to be unusable, and plasma spectroscopy again becomes a diagnostic as well as a means of studying radiation transport and plasma dynamics [19, 20].

In astrophysics, spectroscopic methods remain the only ones available. The study of astrophysical plasmas has benefitted greatly from the 'calibration' of spectroscopic techniques in laboratory plasmas, as well as from the greatly improved and increased data on oscillator strengths and collisional cross-sections stemming from laboratory work [1–3, 23].

The next two sections describe the most widely used spectroscopic methods for measurement of temperature and electron density. The determination of astrophysical abundances, although a very important aspect of spectrophysics, follows so closely from the absorption and emission measurements described in the previous chapter, once T and, where relevant, N_e are known, that further discussion here seems unwarranted. All the same, it is worth emphasizing the dependence of abundance measurements on reliable absolute oscillator strengths; quite drastic changes in abundance estimates have not infrequently followed new determinations of f-values, even for such long-studied and abundant elements as iron.

A cautionary note must be attached to the word 'temperature'. The values of T obtained by different methods – velocity distribution among ions, say, compared with population distribution among excited states – can agree only if LTE is a good approximation. This is not to say that useful measurements cannot be made on non-LTE plasmas. The intensity ratio of two lines from different initial

levels gives directly the population ratio for those two levels if the relative *f*-values are known, and although this population ratio may not define a temperature it can still give useful information – a value for N_e, for example, if the collisional cross-sections are known.

13.15 Spectroscopic measurement of temperature

All but the first and last methods described here measure, by one means or another, the populations of excited states of atoms or ions; they therefore lead in the case of LTE to a 'Boltzmann' or 'Saha' temperature. The first and last methods measure velocity distributions of atoms or ions and electrons, respectively, and give, if these distributions are properly Maxwellian, a gas kinetic temperature and an electron temperature.

13.15.1 GAS KINETIC TEMPERATURE

It is shown in section 10.3 that the FWHM of a Doppler broadened line is determined entirely by the atomic or molecular weight of the atom or ion and the temperature parameter describing its velocity distribution. By equation 10.17,

$$\frac{\delta\nu_D}{\nu} = \frac{\delta\lambda_D}{\lambda} = \frac{2}{c}\sqrt{\left(\frac{2RT\ln 2}{M}\right)} = 7.16 \times 10^{-7}\sqrt{\left(\frac{T}{M}\right)}$$

When using this relation to determine T, it is essential to make sure that instrumental width and pressure broadening have been allowed for properly. As pressure effects usually show up first in the line wings, the FWHM may not be greatly affected, and it may be possible to avoid the effects altogether by using a transition between inner shells, sufficiently shielded by outer electrons to have a pure Doppler profile. In certain cases Doppler broadening may be caused by bulk motions of the gas – macroscopic turbulence or ions in an electric field, for example. Non-random velocities of this kind do not define a temperature. Finally, as hinted at in section 10.9, Doppler and pressure broadening are not always completely independent of one another, and correlation effects are occasionally significant.

13.15.2 POPULATION OR EXCITATION TEMPERATURE

The temperature parameter in the Boltzmann distribution is usually found by measuring relative intensities of emission lines. The methods and problems are basically the same as those described in section 12.3 for the measurement of *f*-values. (Absorption measurements are possible in principle, but weak absorption lines are more difficult to measure accurately than weak emission lines because the signal-to-noise ratio is less favourable.)

The usual practice is to avoid the hazards associated with absolute intensity measurements by working with relative intensities. Equation 11.47 can conveniently be written

$$I_{21} = \text{const. } N_2 A_{21} \nu_{12} = \text{const. } g_2 A_{21} \nu_{12} e^{-E_2/kT}$$

A plot of $\log (I\lambda/gA)$ against E for several spectral lines should be a straight line of slope $1/kT$, with deviations from the Boltzmann distribution, or from optical thinness, showing up as points off the line. The main errors associated with this method are inaccuracies in the A-values and the limited range of E-values that can be used. It is usually difficult to find suitable lines with a wide spread of upper states, and if they can be found they tend to fall in very different spectral regions, making the intensity comparison difficult.

Vibrational and rotational levels of molecules give useful Boltzmann plots for temperatures below about 5000 K. They have the advantage that their relative transition probabilities can usually be calculated rather accurately. The relative rotational line strengths are the Hönl–London factors $s_{J'J''}$ discussed in section 11.10. In terms of these factors and equation 11.65 relating transition probability to line strength we have

$$I = \text{const. } (2J'+1)\nu^3 \frac{S_{J'J''}}{2J'+1} \nu e^{-E_{J'}/kT}$$

leading to

$$\log\left(\frac{I\lambda^4}{S_{J'J''}}\right) = \text{const.} - \frac{E_{J'}}{kT} = \text{const.} - \frac{BJ'(J'+1)}{kT}$$

A plot of $\log (I\lambda^4/S)$ against $J'(J'+1)$ should give a straight line with a slope of B/kT. A similar expression should hold for integrated band intensities, replacing E_J by E_v and the Hönl–London factors by the relative band strengths. The approximation made in using Franck–Condon factors for these band strengths (section 11.10) usually renders this method rather less reliable. A modification for rotational lines is based on band maxima: as J increases the intensity in any branch of a band first increases because of the Hönl–London factor, which is approximately proportional to $2J'+1$, and then decreases because of the exponential factor. For two lines of equal intensity on either side of the maximum, J_a and J_b, it follows from the last equation that

$$\log(S_a \nu_a^4) - \log(S_b \nu_b^4) = \frac{B}{kT}[J_a(J_a+1) - J_b(J_b+1)]$$

from which T can be found by matching rather than measuring intensities. Since the lines have the same intensity and approximately the same frequency, the method is applicable even when the plasma is not truly optically thin, but it is not easy to check its accuracy. The rotational lines of certain molecules whose spectra are nearly always observed as impurities in arcs and flames, notably OH and CN, are frequently used for thermometric purposes. High resolution is

required to resolve the rotational lines, especially if the molecule is not a hydride.

13.15.3 IONIZATION TEMPERATURE

A much wider effective range of upper levels can be obtained by using different stages of ionization, together with Saha's equation to relate the number densities. The relative populations of excited levels of different stages of ionization are given by the combined Saha–Boltzmann equation (equation 13.23). In an optically thin plasma the intensity of a line $I_{i,k}$ from the kth level of the ion relative to that of a line from the jth level of the atom $I_{a,j}$ is given by

$$\frac{I_{i,k}}{I_{a,j}} = \frac{A_{i,k}}{A_{a,j}} \frac{2g_{i,k}}{g_{a,j}} \frac{\lambda_{i,k}}{\lambda_{a,j}} \frac{(2\pi mkT)^{3/2}}{h^3} \frac{1}{N_e} e^{-(\chi + E_k - E_j)/kT}$$

Note that one has to know N_e as well as the A-values. Obviously the same method can be used for any two adjacent stages of ionization.

It is possible to assess the spatial or temporal variation of T from the appearance of different stages of ionization, on the basis that, as shown in section 13.5, one stage of ionization dominates all others at any given temperature. If the intensity of one suitable line from each stage is calculated and the result plotted as a function of T, as shown in Fig. 13.12, temperature changes can be estimated by comparing measured relative intensities with the plot. 'Estimated' is likely to be a better description than 'measured' for this method.

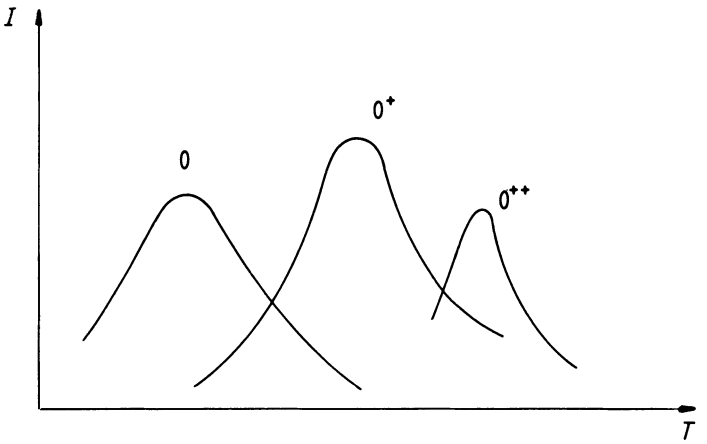

Figure 13.12 Intensities of lines from successive stages of ionization of oxygen as a function of temperature.

13.15.4 REVERSAL MEASUREMENTS

The reversal method much used for flame temperatures is a particularly accurate way of combining emission and absorption methods by comparing directly the populations of the upper and lower states of a transition. If a given spectral line is observed against a background source of variable radiance, such as a filament lamp, the line appears bright against the background when the lamp is dim. As the lamp is turned up, the line disappears altogether and then reappears as a dark line against a bright background. At the disappearance point the brightness temperature of the lamp is equal to the excitation temperature of the flame for that particular transition. This can be shown as follows. Equation 11.52 gave for the radiance of a homogeneous slab of hot gas of thickness l

$$L_\nu(l) = L^{\mathrm{B}}(\nu, T)\,(1 - \mathrm{e}^{-k'_\nu l})$$

If the same slab is illuminated by a background source of radiance $L'(\nu, T')$, the ratio of transmitted to incident radiance is

$$\frac{L'(l)}{L'(0)} = \frac{I_\nu(l)}{I_\nu(0)} = \mathrm{e}^{-k'_\nu l}$$

The power absorbed (per unit area per steradian) is therefore

$$L'(0)\,(1 - \mathrm{e}^{-k'_\nu l})$$

At the reversal point this must equal what is put back by the gas, giving

$$L'(\nu, T') = L^{\mathrm{B}}(\nu, T)$$

If the background source is itself a blackbody, T and T' must be the same. It is apparent from the derivation of equation 11.52 in section 11.9 that T is the Boltzmann temperature for the two states concerned.

The reversal method in its direct form is limited to plasmas of relatively low temperature, for which a suitable adjustable background is available. It can be extended to much hotter plasmas by comparing the total emission with the equivalent width, using a background source of fixed temperature for the latter. In the absence of this background source the radiance integrated over the line profile is

$$L_{\mathrm{p}} = \int_{\text{line}} L_\nu \,\mathrm{d}\nu = L^{\mathrm{B}}(\nu, T) \int_{\text{line}} (1 - \mathrm{e}^{-k'_\nu l})\,\mathrm{d}\nu$$

or

$$L^{\mathrm{W}}(\nu, T) \int_{\text{line}} (1 - \mathrm{e}^{-k_\nu l})\,\mathrm{d}\nu$$

(assuming Wien's approximation to be valid). In absorption against a background source of much greater radiance L' the equivalent width of the line is

given by (equation 11.42)

$$W_\nu = \int_{\text{line}} (1 - e^{-k_\nu l})\, d\nu$$

If Wien's approximation is valid,

$$\frac{L_p}{W_\nu} = L^W(\nu, T) = \frac{2h}{c^3}\nu^3 e^{-h\nu/kT}$$

The procedure is to measure L_p and W_ν for several lines; a plot of $\log[L_p/(W_\nu \nu^3)]$ against ν should give a straight line of slope $-h/kT$, with departures from LTE showing up as deviations from the straight line. If L_p is not negligible compared with the background, an appropriate correction to W must be made. The sensitivity of this method, like that of section 13.15.2 depends on the spread of wavelength, but there are two substantial advantages: transition probabilities are not required, and the plasma need not be optically thin.

13.15.5 OPTICALLY THICK LIMIT

As shown in section 11.9, the peak height of any emission line increases with increasing optical depth until it hits the blackbody ceiling as $k_\nu l \to \infty$ – see also equation 13.54. In principle, therefore, the temperature of an optically thick plasma in LTE can be found from an absolute intensity measurement. Whether or not the line is optically thick can usually be checked by attempting to increase the intensity by reflecting the radiation back through the plasma or by viewing it end-on instead of side-on, but it is usually necessary to use sufficient spectral resolution to pick out the centre of the line. The method is inherently rather inaccurate because for $T > 10\,000$ K the peak of the blackbody distribution is in the ultra-violet where absolute intensity measurements are difficult, and in the visible region where the measurements are easier the intensity varies rather slowly with T.

An alternative is to measure the absolute intensity of the continuous emission at a wavelength sufficiently long that the plasma is optically thick (section 13.12 and Fig. 13.11). This is feasible only if the optically thick limit is reached at a frequency below the plasma frequency. By equation 13.48, $k_\nu l$ increases with $N_e N_i l/\sqrt{T}$ – i.e. with $N_e^2 l/\sqrt{T}$ – whereas the plasma frequency ν_p increases with $\sqrt{N_e}$, so for a given T and l the method cannot be used below a critical electron density. Figure 13.13 shows the approach to optical thickness and the corresponding plasma frequency at three different electron densities for a hydrogen plasma 1 cm thick at 100 000 K. The critical density in this case is about 10^{17} cm^{-3}.

13.15.6 INTENSITY OF CONTINUOUS RADIATION

Equations 13.51 and 13.52 for the bound–free and free–free emission of an

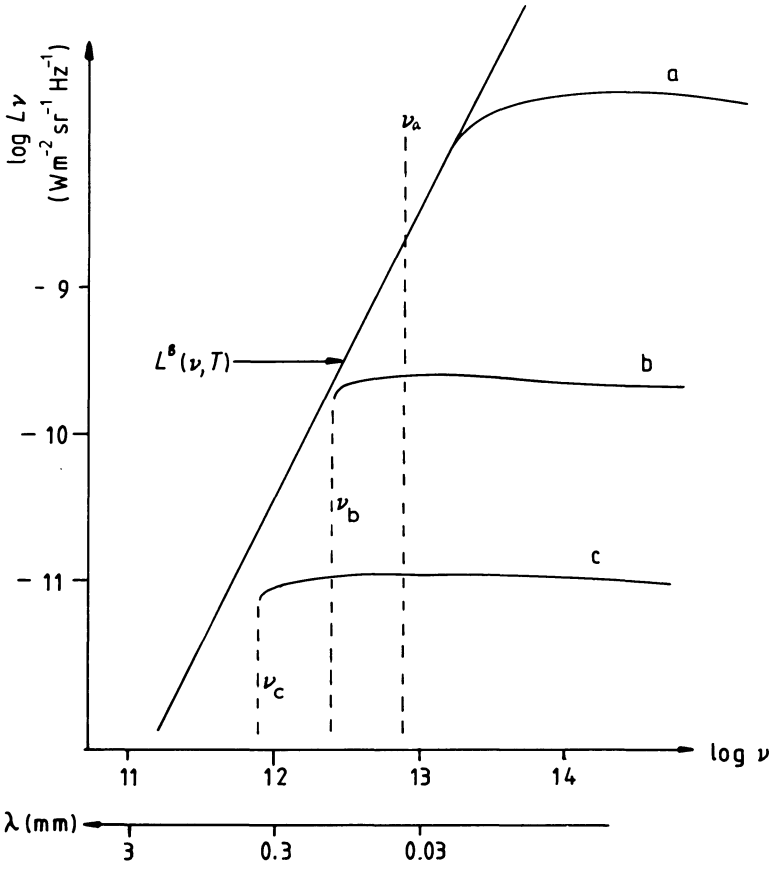

Figure 13.13 Approach to blackbody radiation at different electron densities. The curves show the radiance of a hydrogen plasma of depth 1 cm at 10^5 K as a function of frequency for N_e values of (a) 10^{18} cm^{-3}, or 10^{24} m^{-3}, (b) 10^{17} cm^{-3}, and (c) 10^{16} cm^{-3}. The vertical broken lines are the corresponding plasma frequencies, and the solid line is the blackbody emission for 10^5 K (Rayleigh–Jeans region).

optically *thin* hydrogen-like plasma show that the frequency dependence is governed by the term $\exp(-h\nu/kT)$. In the regions between the absorption edges a plot of $\log j_\nu$ versus ν should give a straight line of slope h/kT, where T pertains to the Maxwellian velocity distribution of the electrons and not to any population ratios.

An alternative for plasmas for which the absolute value of the continuous emission can be calculated with sufficient assurance is to use the ratio of the intensity of the continuum to that of a neutral emission line. The first depends on $N_e N_i$ and the second on N_a, and in principle this allows T to be found from Saha's equation on the assumption of partial LTE.

13.16 Spectroscopic determination of electron density

13.16.1 LINE BROADENING

If inelastic collisions can be ignored, line broadening by charged particles can be ascribed to the linear Stark effect (C_2/r^2) for hydrogen and to the quadratic Stark effect (C_4/r^4) for other atoms. The FWHM of the line depends on N_e. Such measurements are made whenever possible on H-atoms or H-like ions because of the greater broadening and the greater accuracy with which it can be calculated. The hydrogen lines are well described by a quasi-static model in which

$$\delta\lambda = C(N_e, T)\, N_e^{2/3}$$

where C depends only weakly on N_e and T. For H_β, the second line of the Balmer series and the most overworked line in the spectrum for this purpose, $\delta\lambda \simeq 1$ nm for $N_e \simeq 10^{16}$ cm^{-3}, and the determination of N_e should be good to about 5%.

Atoms that are not H-like are subject to the quadratic effect. The broadening is predominantly due to electron impacts, with a FWHM proportional to N_e. Calculations of the constant of proportionality are usually subject to some uncertainty, but a number of lines commonly found in laboratory plasmas have been 'calibrated' in known conditions [10].

13.16.2 ABSOLUTE INTENSITY OF CONTINUOUS RADIATION

Equations 13.51 and 13.52 show that j_ν for both bound–free and free–free emission depends on N_e. Apart from the usual difficulty of making absolute intensity measurements, the calculations are of doubtful reliability except for H-like and He-like plasmas, and the Z^2-factor means that any impurities contribute disproportionately to the intensity. The equations are applicable only in the optically thin region.

13.16.3 PLASMA FREQUENCY

From equation 13.5,

$$\nu_p = \left(\frac{e^2 N_e}{4\pi^2 \varepsilon_0 m}\right)^{1/2} = 9N_e^{1/2} \quad \text{Hz}$$

and since ν_p acts as a cut-off frequency for emission and transmission of radiation N_e can in principle be determined from ν_p, independent of temperature. The cut-off frequency cannot be determined with great precision, and in the case of laboratory plasmas it is usually preferable to measure the refractive index of the plasma at frequencies greater than ν_p by laser interferometry and apply equation 13.11.

13.16.4 INGLIS–TELLER LIMIT

For Stark broadening the quantum number of the last distinguishable line of a series is related to the electron density as described at the end of section 13.6, and a crude estimate of N_e may be made from this relation when the much more reliable laboratory methods cannot be used. The uncertainty relates to determining the point at which the lines actually do merge, particularly as *any* form of broadening contributes. Probably the only real application of this method has been the use of the Inglis–Teller limit for the Balmer series of hydrogen to estimate N_e in certain stellar atmospheres.

References

GENERAL – EQUILIBRIUM, IONIZATION, CONTINUOUS RADIATION, ETC.

1. Aller, L.H. (1963) *The Atmospheres of the Sun and Stars*, Ronald Press, New York.
2. Cowley, C.R. (1970) *The Theory of Stellar Spectra*, Gordon and Breach, New York.
3. Mihalas, D. (1978) *Stellar Atmospheres*, Freeman, San Francisco.

ELECTRON COLLISIONS

4. Bransden, B.H. and Joachain, C.J. (1983) *Physics of Atoms and Molecules*, Longman, London.
5. Hasted, J.B. (1964) *Physics of Atomic Collisions*, Butterworths, London.
6. Massey, H.S.W. and Burhop, E.H.S. (1952) *Electronic and Ionic Impact Phenomena*, Oxford University Press.

NEGATIVE IONS

7. Massey, H.S.W. (1976) *Negative Ions*, 3rd edn, Cambridge University Press.
8. Hotop, H. and Lineberger, W.C. (1985) Binding energies in negative ions, *J. Phys. Chem. Ref. Data*, **14**, 731.

PLASMA SPECTROSCOPY

9. Griem, H.R. (1964) *Plasma Spectroscopy*, McGraw-Hill, New York.
10. Griem, H.R. (1974) *Spectral Line Broadening by Plasmas*, Academic Press, New York.
11. Huddlestone, R.H. and Leonard, S.L. (eds) (1968) *Plasma Diagnostic Techniques*, Academic Press, New York.
12. Lochte-Holtgreven, W. (ed.) (1968) *Plasma Diagnostics*, North-Holland, Amsterdam.
13. Burgess, D.D. (1972) Spectroscopy of laboratory plasmas, *Space Sci. Rev.*, **13**, 493.
14. Cooper, J. (1966) Plasma Spectroscopy, *Rep. Prog. Phys.*, **22**, 35.
15. Key, M.H. and Hutcheon, R.J. (1980) Spectroscopy of laser-produced plasmas, *Adv. Atomic Molec. Phys.*, **16**, 201.

Articles in Vol. 2 of (1984) *Applied Atomic Collision Physics* (Barnett and Harison, eds), Academic Press, New York.

16. McWhirter, R.W.P. and Summers, H.P. 'Atomic radiation from low density plasmas'.

17. Peacock, N.J. 'Diagnostics based on emission spectra'.
18. Weisheit, J.C. 'Atomic phenomena in hot dense plasmas'.

DENSE PLASMAS

19. Burgess, D.D. (1982) Spectroscopy of dense laser-generated plasmas, in *Laser–Plasma Interactions 2* (ed. R.A. Cairns) Scottish Universities Summer School in Physics, Edinburgh.
20. More, R. (1983) Atomic processes in high density plasmas, in *Atomic and Molecular Physics of Controlled Thermonuclear Fusion*, (ed. C.J. Joachain and D.E. Post), Plenum Press, New York.

COLLISIONAL EXCITATION AND IONIZATION RATES

21. Sobelman, I.I., Vainshtein, L.A. and Yukov, E.A. (1981) *Excitation of Atoms and Broadening of Spectral Lines*, Springer-Verlag, Berlin.
22. Bell, K.L., Gilbody, H.B., Hughes, J.G., Kingston, A.E. and Smith, F.J. (1983) Recommended data on the electron impact ionization of light atoms and ions, *J. Phys. Chem. Ref. Data*, **12**, 891.
23. *Review articles on astrophysical applications* may be found in the *Annual Reviews of Astronomy and Astrophysics.*

Appendix

General physical constants

electron charge	e	1.60219×10^{-19} C $(4.80324 \times 10^{-10}$ esu)
electron mass	m	9.10953×10^{-31} kg $(9.10953 \times 10^{-28}$ g)
charge/mass ratio of electron	e/m	1.758805×10^{11} C kg^{-1} $(5.27276 \times 10^{17}$ esu g^{-1})
proton mass	m	1.67265×10^{-27} kg $(1.67265 \times 10^{-24}$ g)
proton/electron mass ratio	m_p/m	1836.15
Planck's constant	h	6.62618×10^{-34} J s $(6.62618 \times 10^{-27}$ erg s)
'h-bar' ($h/2\pi$)	$\hbar$	1.05459×10^{-34} J s
Boltzmann's constant	k	1.3807×10^{-23} J K^{-1} $(1.3807 \times 10^{-16}$ erg K^{-1})
velocity of light	c	2.997925×10^{8} m s^{-1} $(2.997925 \times 10^{10}$ cm s^{-1})
Avogadro's number	N_A	6.02205×10^{26} molecules kmole^{-1} $(6.02205 \times 10^{23}$ molecules mole^{-1})
Loschmidt's number	N_L	2.68678×10^{25} molecules m^{-3} $(2.68678 \times 10^{19}$ molecules cm^{-3})
permittivity of vacuum	ε_0 $1/4\pi\varepsilon_0$	8.85419×10^{-12} C^2 N^{-1} m^{-2} or F m^{-1} 9×10^9 N m^2 C^{-2}
permeability of vacuum	μ_0	$4\pi \times 10^{-7}$ Wb A^{-1} m^{-1} or H m^{-1}

Spectroscopic constants

Rydberg constant	$\frac{2\pi^2 m}{100ch^3}\left(\frac{e^2}{4\pi\varepsilon_0}\right)^2$	R_∞ chR_∞	$1.09737312 \times 10^5\ \text{cm}^{-1}$ 13.6058 eV
for H	$\left(m \to \frac{mm_p}{m+m_p}\right)$	R_H	$1.09677580 \times 10^5\ \text{cm}^{-1}$
radius of 1st Bohr orbit	$\frac{\hbar^2}{m}\left(\frac{4\pi\varepsilon_0}{e^2}\right)$	a_0	5.29177×10^{-11} m 0.529177 Å
fine structure constant	$\frac{1}{\hbar c}\left(\frac{e^2}{4\pi\varepsilon_0}\right)$	α $1/\alpha$	7.29735×10^{-3} 137.0360
Bohr magneton	$\frac{e\hbar}{2m}$	μ_B	$9.274096 \times 10^{-24}\ \text{J T}^{-1}$ $0.46686\ \text{cm}^{-1}\ \text{T}^{-1}$
Bohr magneton (CGS units)	$\frac{e\hbar}{2mc}$	μ_B	$4.6686 \times 10^{-5}\ \text{cm}^{-1}\ \text{G}^{-1}$ $9.274096 \times 10^{-21}\ \text{erg G}^{-1}$
nuclear magneton	$\frac{e\hbar}{2m_p}$	μ_N	$5.05082 \times 10^{-27}\ \text{J T}^{-1}$

Note: For CGS formulae for R, a_0 and α, replace $e^2/4\pi\varepsilon_0$ with e^2.

Energy conversion

$1\ \text{J}\ (10^7\ \text{erg}) = 6.242 \times 10^{18}\ \text{eV} \Rightarrow 5.034 \times 10^{22}\ \text{cm}^{-1}$

$1\ \text{eV} = 1.602 \times 10^{-19}\ \text{J} \Rightarrow 8065\ \text{cm}^{-1}$

$1\ \text{cm}^{-1} \Rightarrow 1.986 \times 10^{-23}\ \text{J} = 1.240 \times 10^{-4}\ \text{eV}$

$k = 1.380 \times 10^{-23}\ \text{J K}^{-1} = 0.861 \times 10^{-4}\ \text{eV K}^{-1} \Rightarrow 0.695\ \text{cm}^{-1}\ \text{K}^{-1}$

$hc = 1.98648 \times 10^{-23}\ \text{J cm}^{-1} = 1.23985 \times 10^{-6}\ \text{eV m}^{-1}$

Index

Absorption
  by atmospheric gases 14, 15, 167
  coefficient 256, 293–6, 300
    atomic 295
  continuous 360–6
    cross-sections 361–2, 365
    molecular 87–9
    see also Photo-ionization
  integrated 290, 293, 295, 299, 323–4
  probability coefficient, *see* Einstein coefficients
  spectroscopy 11, 109–13
    background sources for 110–13
    microwave 229
  total *see* Equivalent width
Abundances 307
A. C. Stark effect 221–2
Adiabatic approximation 272
  breakdown of 276–7
Airy distribution *see* Intensity distribution of Fabry–Perot
Aliasing 194–5
Alternation of intensities 86–7
Ångstrom unit 4, 5
Angular momentum
  orbital 25
  rotational 75
  spin 27, 59
  total 38, 40, 59, 77
Anharmonicity constant, x_e 74
'Anomalous' terms 9
Anti-bonding orbitals 68–9
Anti-Stokes lines 248–9, 250
Anti-symmetric functions 36–7, 86–7
Apodization 192
Approximations 8
Arcs 105–6
  argon mini- 99
  stabilized 106, 322
Array detectors 121–2
Astigmatism of concave grating 156, 157
Astrophysics 2, 46, 62, 307
Atomic beam magnetic resonance 234–42
Atomic clock 16, 242
Atomic energy levels, tables of 9
Atomic units of line strength 292–3
Autler-Townes effect, *see* A. C. Stark effect
Auto-ionization 55–8

Balmer 2
Balmer series 9, 10
Band origin 80, 83
Band oscillator strength *see* Oscillator strengths
Band strength 314
Beam foil source 109
  for lifetime measurements 331–2
Beam-splitter 171–2, 188, 190, 198
Binary approximation 266, 272
Black body 99
  radiation 96–8
Blaze, *see* Grating
Bohr 2
Bohr magneton 38
Bohr model of H atom 8, 20
Bohr orbits 21

Bolometers 116–17
Boltzmann plot 371
Boltzmann relation 100, 286–7, 348
Boltzmann–Saha equation 350, 372
Bond lengths 230
Bonding orbitals 68–9
Born approximation 354
Born–Öppenheimer approximation 66–7
Bound-free transitions 8, 360–1, *see also* Photo-ionization; Radiative recombination
Branching ratios 322, 335–6
Bremsstrahlung 112, 361, 367, 368, *see also* Free–free transitions
Brightness 93
Broadening, homogeneous and inhomogeneous 211

CARS (coherent anti-Stokes Raman spectroscopy) 251–3
Cascading 331
Cavity, laser 204–5
  tuning of 207
Central field approximation 29–31
Centrifugal constant D 76–7
Classical damping constant 259–61
Classical dipole, equivalent 291–2
Classical oscillator 259–61, 297–8
Classical path assumption 272
Coherence narrowing 245
Collisional broadening *see* Pressure broadening
Collisional coefficients 354, 356–7
Collisional cross-sections
  excitation and de-excitation 353–4
  ionization and recombination 353, 359–60
Collisions, inelastic 276–7, 331
Compensating plate 188–9
Condensing lens 134–5
Configuration, electron, definition of 28
Configuration mixing 54–6
Continuous absorption and emission *see* Absorption *and* Emission
Constant deviation spectroscope 136–7
Contrast of Fabry–Perot fringes 177
Contrast of photographic emulsion 119–20
Conversion factors for energy 4, 380
Conversion factors for intensity parameters 290–3
Convolution theorem 192
Core electrons 30, 31
Corona, solar 63, 343, 358
Coronal equilibrium 358–9
Covalent bonding 69
Cross-dispersion
  for echelle 166
  for Fabry–Perot 182–3
Crystal fields 63, 233
Curve of growth 304–8
Curvature
  of spectral lines 142, 156
  of field 142
Cyclotron radiation 368
Czerny–Turner mounting 157

d-electron 28
D terms 34
Δ terms 70
Damping ratio 282
Dark current 118
Debye radius 341–3
Debye sphere 342
Degeneracy 27–8, 246, 289
  see also Statistical weight
Delayed coincidence method (of lifetime measurement) 329–30
De-mixing 322
De-polarization of resonance radiation 243–5
Detachment energy 368
Detailed balance 102, 355
Detectors 11–13, 113–22
  for the infra-red 115–18
  for the ultra-violet 118–19
  photon 113, 115–19
  thermal 113, 115–17
Deuterium lamp 99
Deviation, minimum 137, 140
Diamagnetic shifts 234
Diatomic molecules, constants for 8, 74, 77
Dielectronic recombination 57–8
Diffraction, Fraunhofer 127–9
Dirac comb 194

Dirac, relativistic theory 28
Dispersers 11–13
Dispersion, angular 126
  of grating 150–1
  of prism 137–9
Dispersion, linear 126
  of grating instrument 151
  of prism instrument 139
Dispersion of prism material 139
Dispersion, reciprocal 126
  of grating instrument 151, 156
  of prism instrument 139
Dissociation energy 65–6
Dissociation equilibrium 349
Doppler broadening 262–5
  in beam foil source 331–2
Doppler-free spectroscopy 218–21
Doppler profile 263–4, 306
Doppler shift 263
Doppler width *see* FWHM
Double etalon 185
Double resonance spectroscopy 342–5
Doublet structure 39, 41

Eagle mounting 158, 160, 162
Ebert mounting 157, 159
Echelette gratings 167
Echelle gratings 165–6
Eigenfunctions, -values, definitions of 22
Einstein coefficients 101–4, 288–93, 356–7
Electric dipole line strength 44, 289
Electric dipole moment, permanent 69, 84
Electric dipole transitions 43–4, 241, 244, 288–9
Electric dipole transition moment *see* Transition moments
Electric quadrupole interaction 60–1, 230–31
Electric quadrupole line strength 292–3
Electric quadrupole transitions 45–6, 292
Electrodeless discharge 106–7
Electromagnetic spectrum 4–6
Electron density, spectroscopic determination of 376–7
Electron pressure 351
Electron spin 27
Electron spin resonance (ESR) 233–4
Electron synchrotron *see* Synchrotron radiation
Electron temperature 100, 356–7, 375
Electron volt 4, 5, 380
Electronic energy of diatomic molecules 67–71
Electronic spectra of diatomic molecules 80–3
Emission coefficient, line 308
Emission, continuous 360–2, 366–7
  cross-section 361–2, 367
  molecular 87–9
  of inert gases 89, 111
  of molecular hydrogen 89, 111
Emissivity 98
Energy density of radiation 95, 96
Energy level diagrams
  compilations of 8
  definition of 7
Energy levels, data on 8
Energy levels of H atom 7, 8, 20
Equivalent electrons 38
Equivalent width 302–4, 324–5
Etalon 175
Étendue, *see* Throughput
Exchange integral 36
Exchange interaction 35–7

F, *see* Quantum numbers
f-electron 28
f-sum rule 301–2
f-value, *see* Oscillator strength
Fabry–Perot interferometer 171–2, 175–85
  photographic use 183
  scanning 173, 183–5
Fano line profiles 57
Fastie mounting 157
Fellgett advantage, *see* Multiplex advantage
FFT (fast Fourier transform) 196
Field-independent transition 239, 242
Field ionization 217
Fine structure 38–41
  of H atom 41–2
Finesse 179, 182
Flames 105
Flash photolysis and pyrolysis 109

Flash tubes 110, 111
Flatness of interferometer plates 182
Flux density, definition 93
Forbidden transitions 46, 279, 359
Fourier transforms 187, 193–4
  complex 197
Fourier transform spectroscopy 185–98
  dispersive 197
Franck–Condon factors 313–14
Franck–Condon principle 82, 85
Fraunhofer 1
Fraunhofer lines 1, 2
Free–bound transitions *see* Bound-free
Free–free transitions 8, 360–1, 365, 366–7
Free spectral range
  in Fourier transform spectroscopy 194–5
  of echelle 166
  of Fabry–Perot 178, 182
Frequency doubling and mixing 208–9
Frequency standards 16, 214
Furnaces 109–10
FWHM (full width half maximum), definition 256
  collisional 275
  Doppler (Gaussian) 263
  Lorentzian 261
  natural 258, 261–2
  of diffraction distribution 131
  of Fabry–Perot fringes 178–9
  of Fourier transform instrument function 191–2

g-factor for fine structure *see* Landé g-factor,
  for hyperfine structure *see* Hyperfine structure g-factor
  of free electron 38
  of nucleus *see* Nuclear g-factor
Gaunt factors 363–5
Gaussian distribution *see* Doppler profile
Gerade and ungerade states 71
Ghosts 161, 198
Globar 110
Glow discharges 106–7
Golay cell 116, 117–18
Grain of photographic emulsion 121
Gratings
  blazed 162–3
  concave 153–6
  errors in 161–2
  interferometric (holographic) 164–5
  plane 144–6
  replication of 161–2
  ruling of 161
Grating mountings 156–61
Grazing incidence 168
  mounting 169
Grey body 98
Gyromagnetic ratio 38

Half-value width, *see* FWHM
Hamiltonian operator 21–2
Hanle effect 333–5
Heat pipe 110
Heitler–London molecular wave functions 69
Hole-burning 211–12
Hollow cathode discharge 107
Holtsmark theory of pressure broadening 267–8
Hönl–London factors 312–13
Hook technique 325–8
Hund's coupling cases 77
Hydrogen atom
  Bohr model 20–1
  degeneracy 27–8, 42
  energy level diagram 7
  fine structure 42–3, 220
  hyperfine structure 242
  line series 9
  linear Stark effect 53–4
  wave functions 23–8
Hydrogen ion, negative 368
Hyperfine structure 58–61, 241
  g-factor 239
  in Hanle effect 335
  in level-crossing experiments 246–7
  in magnetic field, *see* Zeeman effect of hfs

I, *see* Quantum numbers
Illumination of spectrograph slit 133–4, 135
Image defects 142

Impact approximation 265, 272–7
Impact parameter 266
Imprisonment of radiation *see* Radiation trapping
Induced emission *see* Stimulated emission
Inductively coupled plasma (ICP) 108–9
Infra-red grating spectrometers 167
Infra-red region, characteristics of 5, 6, 14
Inglis–Teller limit 352, 377
Inner shell excitation 6, 63
Instrument function 126
  Fabry–Perot 178
  Fourier transform 190–3
  grating and prism 131
Intensity, definition of 94
Intensity distribution
  from Fabry–Perot 175–80
  from grating 146–50
  from Michelson interferometer 186
  from slit 131–2
Intensity standards 99
Interferogram 185–9
  symmetry of 197
Interatomic potentials 65–6, 267, 277–9
Inter-nuclear separation 65–6
Inverse auto-ionization *see* Dielectronic recombination
Inversion spectrum 231
Ion traps 223–4
Ionic bonding 69
Ionic energy levels, tables of 9
Ionic spectra 62
Ionization, degrees of 101, 351
Ionization energy 8, 351
  depression of 351–2
Ionization equilibrium *see* Saha's equation
Ions, negative 368
Irradiance, definition of 93
Iso-electronic sequences 62
Isotope separation 217
Isotope shift 61–2
Isotropic radiation 94

J and j *see* Quantum numbers
j–j coupling 41

Jacquinot advantage *see* Throughput advantage

K shell 63
Kayser 4
Kirchhoff's law 256, 309
Kuhn model for pressure broadening 278

L and l *see* Quantum numbers
L–S coupling 41
L shell 63
Lamb and Retherford 231
Lamb dip 213–14
Lamb shift 43, 220, 231
Lambda-doubling 71
Landé g-factor 48
  in Hanle effect 335
Landé interval rule 40, 59
Laporte rule 45
Laser
  cooling 223
  mode structure of 205
  spectral resolution of 212
  stabilization of 213–14
  temporal resolution of 214–15
Lasers
  diode 208
  dye 206–7
  excimer 208
  F-centre 208
  fixed frequency 205–6
  helium-neon 203–4, 206
  ion (argon and krypton) 206
  neodymium 206
  nitrogen 206
  ruby 203–4, 206
  spin-flip 208
Laser-produced plasmas 108, 111
Legendre polynomials 23.
Lennard–Jones potential 271
Lens aberrations 142
Level-crossing spectroscopy 245–7
Lifetimes of excited states 258
Lifetime measurements 328–35
Light-gathering power *see* Throughput
Lindholm model for pressure broadening 274–6

Line series, definition of 9
Line strength 288–93
  conversion factors for 290–3
  of rotational line 314
  see also Electric dipole; Electric quadrupole; Magnetic dipole
Linear combination of atomic orbitals (LCAO) 67–8
Littrow mounting
  for plane grating 157, 158
  for prism 136
Local thermodynamic equilibrium *see* LTE
Lockyer 1
Lorentz theory of pressure broadening 265–6
Lorentzian profile 261, 299, 306
LTE 100, 355–8
  partial 357
Lyman region 14
Lyman series 9, 10

M and m *see* Quantum numbers
M shell 63
Mach–Zehnder interferometer 325
Magnetic dipole
  line strength 292–3
  transitions 45–6, 233, 241, 244
Magnetic dipole moment
  of electron 38
  of nucleus 59
Magnetic resonance spectroscopy 332–42
Magnetic sub-levels 47–50
Masers 202
Mass shift 61
Maxwell velocity distribution 263, 348
Metastable levels 46
Metre, standard 16–17, 185
Michelson interferometer 171–2
  as scanning spectrometer 173, 185–90
Microwave region 5, 6, 15
Microwave spectroscopy 228–31
Millikayser (mK) 4
Milne relations 362
Missing rotational lines 85–6
Molecular beam magnetic resonance 237
Molecular orbitals 67–8
Moments of inertia of molecules 74, 91
Monochromator 12, 13, 136–7
Morse potential 71–2
Moseley's law 63
Mountings *see under* Grating *and* Prism
Multi-photon excitation and ionization 216–17
Multiplets 46, 47
Multiplex advantage 198–200
Multiplex spectroscopy 188
Multiplicity 40

n *see* Quantum numbers
Natural broadening 257–62
  of auto-ionizing levels 57
Natural width *see* FWHM
Nernst glower 110
Noise 113–14
  amplifier 113
  background 113
  detector 114, 199
  photon 113, 200
Noise equivalent power 114
Normal modes 91
Normalization of wave functions 22
Nuclear electric quadrupole moment 60–1, 231, 242
Nuclear g-factor 59, 234, 241
Nuclear magnetic resonance (NMR) 234
Nuclear magneton 59
Nuclear spin 59, 241
Nyquist *see* Sampling theorem

Optical depth 257, 293, 302–4
Optical path difference 126–7, 172–3
Optical pumping 209–11, 245
Optically thick condition 302, 311
Optically thin condition 256, 295, 302
Order number
  of Fabry–Perot 178–9
  of grating 144–6
Orthogonality of wave functions 22
Oscillator strength
  absolute 319
  conversion factors for 290, 291, 293
  definition of 296
  in diatomic molecules 311–16
  'electronic' 315
  of molecular band 314–15

Oscillator strength *contd.*
relative 319, 336–7
of rotational line 314–15
see also Transition probabilities
Overlap integral 85, 313
Overlapping lines 268
Overlapping orders 146, 166

P branch 79–80
p-electrons 28
P terms 34
Π terms 70
Paramagnetic resonance *see* Electron Spin Resonance
Parity 45
Partition function 287–8
divergence of 352
of free particle 348
Paschen–Back effect 49
Paschen–Runge mounting 158, 160, 162
Pauli exclusion principle 27–9, 69
Perturbation theory 31–3
Perturbations
by electric field 51–4
of degenerate states 53–4
phase change from 268, 272–6
time-dependent 33, 43
time-independent 31–3
Phase shift method (of lifetime measurement) 329–30
Photoconductors 115–17
Photodiodes 118
Photographic emulsions 119–21
Photo-ionization 9, 360–65
cross-sections 361, 363
detectors 119
Photomultipliers 118–19
Photon noise *see* Noise
Pinch discharges 108
Planck distribution *see* Blackbody radiation
Plasma, definition of 340
neutrality of 340
Plasma focus 108
Plasma frequency 343–6, 376
Plasma, laser-generated 108, 111
Plasma oscillations 344–7
Plate tilt 139–40
Pockel cell 214
Polarizability
in CARS 251–2
in Rayleigh and Raman scattering 247–8
in Stark effect 51
non-linear 208–9
of perturber 270
Polarization of resonance radiation 242–5
of Zeeman components 48, 240–1
rules for dipole transitions 241
Population inversion 103, 203–4
Power broadening
laser 221
microwave 229
Predissociation 89–90
Pressure broadening 265–81
in microwave spectroscopy 229
Pressure shift 275
Prism materials 139, 141–2
Prism spectrographs and spectrometers 134–43
Pyroelectric detectors 117

Q branch 83
Q-switching 214
Quantum beats 247
Quantum efficiency (of detectors) 118–19, 122
Quantum number F 59, I 59, j 38–9, J 40, J (rotational) 75, K 77, l 23, L 34–5, M_F 239, M_I 238, m_j 39–40, M_J 48, M (rotational) 75, m_l 23, M_L 34, m_s 27, M_S 37, n 23, N 77, S 34–5, v 73, Λ 70, Σ 70, Ω 70
Quasi-static approximation 267–8, 277–9

R branch 79–80
Rabi frequency 221–2
Radial charge density 27, 31
Radiance, definitions of 93, 96
Radiation damping 259
Radiation trapping 331, 335
Radiative recombination 9, 360–3, 366
Radiative transfer 308–11

Radiofrequency region 5, 6, 15
Raman spectroscopy 83, 247–53
  cross-sections 250
Ramsey fringes 236
  optical 220–1
Rayleigh scattering 247–8
Rayleigh's criterion 129–30, 180–1
Rayleigh–Jeans formula 97
Reciprocity law 120
Recombination *see* Collisional coefficients *or* Radiative recombination
Reflectivities 168
  of Fabry–Perot coatings 175, 179, 182
Refractive index 299–300, 347
  classical theory of 297–8
  complex 298
  of free electron gas 345
  of prism materials 139
Resolution limit 127
  of grating 151
  of Fabry–Perot 179, 180–1
  of Fourier transform spectrometer 191, 195–6
Resolving power 126–7
  in microwave spectroscopy 229
  of prism 140–1
  of grating 151
  of echelle 165
  of Fabry–Perot 174, 179
  of Fourier transform spectrometer 174, 191
  plate-limited 152
  slit-limited 152–3
Resonance broadening 269, 272
Response curve of photographic emulsion 120
Responsivity of detector 114
Retroreflectors 189
Ritz combination principle 16
Rotational constant B 75, 230
Rotational energy of diatomic molecules 74–7
Rotational spectra 77–8
Rotational wave functions 75
Rowland circle 153–5
Russell–Saunders coupling *see* L–S coupling
Rydberg 2
Rydberg constant 8, 20–1, 220
Rydberg states 218

s-electrons 28, 30
S terms 34
Σ terms 70
Saha's equation 101, 349–51, 359, 372
  for negative ions 368
Sampling errors 198
  interval 193–5
  theorem 195
Satellites 278–9, 281
Saturation
  of absorption 257
  of emission 257, 311
  in Doppler core 307
Saturation spectroscopy 210–11, 218–20
Schrödinger's equation 20–2
Schumann region 14
Schwarzchild factor 121
Selective excitation and ionization 216–17
Selection rules 44
  in magnetic field 48, 240–2, 244
  for electric dipole radiation 45
  for electric quadrupole radiation 45
  for electronic transitions in molecules 80, 83
  for hyperfine structure 60
  for magnetic dipole radiation 45, 233, 240
  for Raman spectroscopy 249
  for rotational transitions 78, 84
  for vibrational transitions 78, 85
Self-absorption 257, 311
Self-consistent field 30, 67
Self-reversal 311
Series limit 9
  depressed 352
Seya mounting 159, 162
Shell structure 27–9
Shift of spectral lines 269, 275, 276, 278
Shift-to-width ratio 276
Shock tubes 107
Signal-to-noise ratio *see* Noise
Singlet terms 35–8
Sirks focus 155, 156

Slit function *see* Instrument function
Slit width 130–2
Sources 11–13, 98–9, 104–13
Source function 256, 309–10
Spark sources 106
Specific detectivity D* 114–17
Spectrochemical analysis 1, 109
Spherical harmonics 23–6, 75
Spherical polar co-ordinates 22
Spin-orbit interaction *see* Fine structure
Spontaneous emission 101–4, 204
Stark broadening 269, 271–2, 376
Stark effect 51–4, 231
State sum *see* Partition function
Statistical weight 100, 287
  of free particle 348
  see also Degeneracy
Stepwise excitation 215
Stimulated emission 102–4, 202
Stokes line 248–50
Suppressed orders 185
Symmetry of wave functions 36–7, 71, 86–7
Synchrotron radiation 99, 112–13

Temperature
  blackbody 96, 374
  brightness 98
  colour 98
  electron 100, 357, 375
  excitation 357, 370
  gas kinetic 370
  ionization 372, 375
  reversal 373
  spectroscopic measurement of 370–5
  thermodynamic 355
Term value, definition of 7
Thermodynamic equilibrium 355 *see also* LTE
Thermopile 116–17
Throughput 94, 132–3
  advantage 198–200
  of grating spectrometer 152–3
  of scanning interferometer 174
Total reflection 168
Transition moment of diatomic molecule 84
Transition probabilities 44
Transition probabilities *contd.*
  accuracy of 319
  bibliography of 318
  calculation of 320–1
  conversion factors for 290–93
  diatomic molecules, for 84, 318
  measurements of
    in absorption 322–5
    in emission 321–2
    by dispersion 325–8
    tables of 318
    two-photon 215
    see also Electric dipole; Magnetic dipole; Electric quadrupole; Oscillator strength; Branching ratios
Triplet terms 35–8
Two-photon transitions 215

Ultra-violet grating spectrometers 167–9
Ultra-violet region, characteristics of 5, 6, 14
Units 17–18, 340

Valence electrons 29, 301
Van der Waals broadening 270–1
Vector model, introduced 34–5
Velocity of light 17, 214
Vibration-rotation spectra 77–80
Vibrational constant ω_e 73
  energy of diatomic molecules 71–4
  frequency, classical 72–3
  wave function 73
Visibility of interference fringes 196
Visible region, defined 5–6
Voigt profile 281–4
  tabulated 282
Volume shift 61–2

Wadsworth mounting
  of grating 158, 161, 162
  of prism 137
Wave functions, definition of 21
Wavelength standards 15–17, 185, 214
Wavelength tables 9–10
Wavenumber, definition 4, 17
Weisskopf theory of pressure broadening 273–4
Weisskopf radius 273–4

Wien's approximation 98
Wien's displacement law 97
Wood's anomalies 163

X-ray region, characteristics of 5, 6
X-ray spectra 63–4
X-unit 4, 5

Z-expansion 320
Zeeman 2
Zeeman effect 47–50
Zeeman effect of hyperfine structure 237–42, 244
Zeeman splittings *see* Magnetic sub-levels
Zero-field level crossing *see* Hanle effect
Zero-point energy 74